THIS WEEK

EDITORIALS

SCHOOLS UK practical changes will leave young scientists unskilled **p.288**

WORLD VIEW Sound science? Colin Macilwain is not sure **p.289**



LIONS Genetic analysis reveals five jungle-king lines **p.290**

The democracy carousel

European law has allowed citizens to force a debate on human embryonic stem cells less than a year after the previous one. This fruitless democratic exercise has left scientists spinning in uncertainty.

Winston Churchill famously said that democracy was the worst form of government, except for all the others. Scientists in the European Union (EU) who work with human embryonic stem cells have more reasons than most to see the systems's flaws. Yet again, they have been forced to discuss and defend the purpose and ethics of their work. Once more, they must watch as possible curbs on their research are proposed.

The trigger this time is the latest in a series of European Citizens' Initiatives (ECIs): petitions that were introduced in 2012 and that automatically prompt a formal public hearing in the European Parliament when they reach more than 1 million signatures across at least seven EU countries. The 'One of Us' petition, which was signed by more than 1.7 million people across all 28 EU countries, calls for a ban on financing any activity that requires the destruction of human embryos, directly or indirectly — so in addition to forbidding EU funding for work on human embryonic stem cells, it proposes a block on aid for agencies that offer abortion advice.

If this topic sounds rather familiar, that's because the same sensitive issue was extensively discussed by the European Parliament in 2006, and again in 2013, before the launch of the multi-billion-euro Seventh Framework and Horizon 2020 research programmes, respectively. In both cases, and after wide consultation, the EU decided to fund such research, provided that approved projects used existing human embryonic stem-cell lines and respected Europe's variable national laws. Yet, in a hearing on 10 April, the European Parliament dutifully rehearsed the same arguments that had led to its previous decisions.

The atmosphere in the crowded auditorium was less decorous than European parliamentarians are used to. Scornful booing erupted, for example, when the parliament's legal-committee representative, Françoise Castex, declared that there was no reason in law to stop funding human embryo research.

The European Commission must prepare a report responding to the One of Us initiative before 28 May, addressing whether any EU legislation could or should be changed in response. It should not.

European scientists have been unsettled by the One of Us initiative, and also by the prospect of a parliamentary hearing of another ECI, 'Stop Vivisection', that calls for the 2010 legislation on the use of animals in research to be replaced by a new directive banning all animal experimentation. That hearing is likely to take place in September.

What is the value of these new efforts in participatory democracy? The general aim sounds noble; it is, of course, good to be able to hold power to account and to involve citizens in setting agendas for discussion. But in practice, and certainly in the case of these two initiatives, they have little democratic merit. The uses of human embryonic stem cells and animals in research have both been discussed very recently. The 2010 animal legislation represented a hard-fought-for compromise that was agreed by EU member states and the European Parliament only after more than a decade of debate involving consultation with all sides. The even-more-recent legislation on funding of research with human embryonic stem cells also represents a compromise in which all sides, including the representatives of One of Us, had their say. To allow a group representing less than 0.4% of the EU's 500 million or so inhabitants to reignite the debate after such a short time seems more of an anti-democratic act than an enlightened one. Worse, the commission's report will of necessity have to repeat the arguments that led to the 2013 decision to fund some research using human embryonic stem cells, again reflecting that a majority supported the compromise. That opens the door to allegations that the EU invites ECIs — and then ignores them.

"What is the value of these new efforts in participatory democracy?"

ECIs are here to stay. The European Commission is quick to point out that they do not represent direct democracy — there is no obligation to change rules in response to them, unlike the recent Swiss referendum curbing immigration, which is now making great difficulties for scientists (see *Nature* **506,** 277; 2014). But they still create an undesirable atmosphere of uncertainty. They mean that researchers will have to increase their efforts to keep the achievements of science in the headlines.

When it comes to complex, highly emotional issues, passionate minority groups can easily and quickly drum up well-supported petitions in a way that scientists cannot (although patient groups could, and perhaps should, think about doing so). Scientists and advocates can, however, build a counterbalance by continuing to present their work as necessary to the well-being of all members of society — however they may vote. ■

Cancer crossroads

Efforts to understand cancer genomes should take on a fresh focus.

Since the discovery of the first cancer-causing genes in the 1970s, researchers have been eager to catalogue the mutations that can cause cancer. Each mutated gene holds the potential to expand our understanding of what causes the disease — and how to treat it.

The latest progress towards that goal was on display last week, when 18,400 people descended on San Diego in California to attend the annual American Association for Cancer Research meeting. Researchers showed how patterns of mutation can be used to track down the agent that caused them — sunlight, for example, leaves a footprint that differs from a cancer-causing viral infection. Another team had

© 2014 Macmillan Publishers Limited. All rights reserved

catalogued cancer-associated mutations in patients with advanced melanoma, hoping to use the information to tailor immune cells to destroy tumours. And promising initial results were unveiled on targeting a protein called IDH2, mutations in which crop up in many different tumour types (see *Nature* **508,** 158–159; 2014).

It has taken a massive effort to make such achievements possible. Seventeen countries have invested in sequencing cancer genomes through the International Cancer Genome Consortium (ICGC), which aims to sequence more than 25,000 samples. The largest and oldest component of that project is The Cancer Genome Atlas (TCGA) at the US National Cancer Institute (NCI) in Bethesda, Maryland, which intends to characterize 10,000 tumours.

TCGA was initially controversial, because researchers worried that the project would direct funds away from grants to individual investigators. Early results — which showed that cancer mutations were much more abundant and diverse than expected — even evoked schadenfreude in some circles (see *Nature* **455,** 148; 2008). Criticism died down as the project bore fruit.

But TCGA is now winding down: the project ceased collecting new tissue samples last December. The ICGC, too, has virtually stopped accepting proposals for new projects. TCGA aims to complete sequencing and further characterization of its cancer samples by the end of the year. After that, a few groups will receive funding to analyse the data for the next two years. But the programme, as it existed, will cease.

Some cancer researchers have advocated that the programme should continue. Stopping now would be premature, their argument goes, because we have yet to achieve a comprehensive catalogue of cancer-causing mutations. A study published earlier this year determined that compiling a list of mutations present in at least 2% of cancers would require sequencing of about 2,000 tumours in each of at least 50 tumour types (M. S. Lawrence *et al. Nature* **505,** 495–501; 2014). For most cancers, we are still far short of that goal.

On a more optimistic note, the end of the older projects should herald a needed transformation in the field. When TCGA and the ICGC were launched, the technology dictated that only fresh tumour samples could be sequenced. That restriction limited researchers' ability to link sequencing data with clinical outcomes because that information might not be available until years after the sample was taken. Also, at that time, oncologists did not take more than one biopsy of a tumour from a patient, which limited studies of how tumours changed during and after treatment, and how metastases differed from primary tumours.

"The end of the older projects should herald a needed transformation in the field."

Those hurdles have now been largely surmounted. Improved techniques allow researchers to sequence DNA collected from tissue preserved in formaldehyde and embedded in paraffin, opening the door to using banked samples with linked clinical data. And although still an uncommon practice, several clinical trials have shown that many patients are willing to submit to extra biopsies.

To replace TCGA, the NCI intends to sequence tumours from patients enlisted in some of its clinical trials. Other teams will no doubt do the same, allowing researchers to learn more about the importance of a given mutation by associating it with the response to therapy or to overall prognosis.

These changes require a new mindset. Clinical researchers will need to change consent forms for donation of tissue samples, to allow the association of clinical data with the sample. They will have to collect their samples using protocols that ensure utility not just for classical pathology but also for sequencing. Data security, always a concern when dealing with patient information, will need to be bolstered.

Nevertheless, to continue the work is a worthy undertaking. The end of TCGA also represents an opportunity for the field to balance its cancer-genomics projects more evenly between cataloguing mutations and studying their functional significance. Functional studies have had short shrift, whereas sequencing — a simple concept, and easier to communicate to policy-makers and the public — has taken the lead. Correcting that imbalance will lead to exciting discoveries for science and for patients. ■

Practical nonsense

Downgrading practical science will impede UK students in the global workplace.

"Equipped with his five senses, man explores the universe around him and calls the adventure Science." So wrote US astronomer Edwin Hubble in 1929, but was he right? How much should science be an exploration of the senses — as well as a test of knowledge and intellectual flexibility? Does a physics student need to peer through a telescope to grasp the enormity of the Universe? Must a potential chemist grapple with the tap of a titration flask to appreciate the subtleties of reaction synthesis? The UK government is about to take a massive — and massively misguided — gamble that they do not.

Education officials in Britain have decided to remove assessed practical work from the landmark A-level qualification, taken by students aged 16–18 and a prerequisite for university. In doing so, the officials and the school science they oversee have taken a huge step backwards. The move could see an entire generation denied the opportunity to develop an interest in the practical experience of doing science.

The Office of Qualifications and Examinations Regulation (Ofqual), which has made the change, says that such fears are overblown. Practical skills will still be tested, it says, and the results presented as a separate pass/fail mark to accompany the existing A-level letter grade. Schools will be inspected. Practical skills, Ofqual promises, will survive.

There are two problems with this. First, as institutions told Ofqual when it floated the idea last year, universities will still focus on the mark — how else will they differentiate between the thousands of applicants? Second, if the result of a practical test is seen as secondary to the overall grade, then schools will be less concerned with it. UK education has become a ruthless marketplace in which schools are judged by how well they can shift students on to the next stage. Anything that interferes with that is unlikely to be a priority. Students, especially those at poorer-performing schools, will simply be offered fewer lessons in practical science.

As John Baruch pointed out in a World View article last month, the UK change comes at a time when other nations, China chief among them, are placing increased emphasis on practical skills in school-leaving exams (see *Nature* **507,** 141; 2014). The United Kingdom is poised to send its science students into the global competition for scientific and technical jobs with one arm tied behind their backs.

Ofqual made its decision in the face of fierce criticism from leading scientists and science advocates. There were certainly problems with testing practical skills through coursework — long viewed by students as the soft underbelly of academic assessment — but, in this case, the proposed solution is worse. As Imran Khan, chief executive of the British Science Association, puts it: "You wouldn't dream of assessing other practical subjects — like languages, music, or design — by a written test alone, and the same should be true of science." We are back to Hubble's five senses, and the need to stimulate and extend them.

Practical experiments teach the reality of science, with all its frustrations and rewards. The real world, after all, does not always proceed smoothly. As the old joke among physics teachers goes: if an experiment smells, it is chemistry; if it moves, it is biology; and if it doesn't work, it is physics. ■

NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

© 2014 Macmillan Publishers Limited. All rights reserved



Beware of backroom deals in the name of 'science'

The term 'sound science' has become Orwellian double-speak for various forms of pro-business spin, says **Colin Macilwain**.

Scientific and environmental groups enjoyed a small victory in Washington DC earlier this year, when an insidious little item called the sound-science bill was removed from farm legislation signed into law by US President Barack Obama on 7 February.

The defeat of the bill was but a minor skirmish, however, in a broader, global war that rumbles on over how science should be used in regulating everything from smoking to pollution.

The term 'sound science' may sound innocuous — comforting, even. Don't be fooled. In policy circles, its use is now pretty-much confined to the determined band of brothers who make their livings trying to roll back government regulation, by fair means or foul.

Many of these people are neo-conservatives — members of the powerful right-wing cult whose founder, former Trotskyist Irving Kristol, adopted various tricks from the revolutionary politics of the far left. One of these was the casual misuse of the word 'sound' to denote colleagues or ideas as being ideologically reliable.

'Sound science' is thus science that big business knows it can trust. In its name, businesses that sell contentious products are working night-and-day to deflect rules and regulations by exploiting a schoolboy image of science to make their case.

And whatever the issue — nuclear power, pesticides, pharmaceuticals, alcohol, oil or coal — they take the basic game plan from the lessons learned by the tobacco industry over the past twenty years.

I don't think many people realize that shares in British American Tobacco (BAT), the world's second-largest tobacco company, have risen tenfold since the year 2000 — while the main UK stock index, the FTSE 100, has stood still. It has been party time for BAT shareholders, even though everyone knows that cigarette smoking kills thousands. Even as cigarettes are being pushed to the edges of social acceptability in the United States and parts of Europe, much larger and more lucrative markets keep opening up.

That is because the tobacco industry has learned to game the regulatory and trade system. Back in 1993, it hijacked the term sound science, when it set up The Advancement of Sound Science Coalition to question early evidence linking passive smoking to cancer. The coalition pioneered an effective approach that abuses science to hold back regulation for long enough for BAT and others to haul in extraordinary profits.

Today, regulatory efforts such as plain packaging and bans on menthol cigarettes are held up by legal challenge, often using language inserted into global trade agreements by industry lobbyists. This creates boundless opportunities for malfeasance. As Nicandro Durante, BAT's chief executive, told the *Daily Telegraph* on 27 February: "There is no scientific evidence that banning menthol is going to bring any real public health benefit to the population."

THE TOBACCO INDUSTRY HAS LEARNED TO GAME THE REGULATORY AND TRADE SYSTEM.

NATURE.COM
Discuss this article online at:
go.nature.com/xn2hkc

Too right, there isn't. Because no-one has ever tried to ban menthol cigarettes for long enough to test it, the scientific evidence cannot exist. It is no wonder that purveyors of everything from booze to nuclear power look at BAT's chutzpah and share price and think: I'll have some of that.

Big business learned some time ago that its besuited representatives were unlikely to win stand-up public fights against environmentalists or health advocates over issues such as tobacco packaging. They might not even win them in cases, such as genetically modified crops, in which they have a respectable claim. It is much more effective to act behind the scenes and cut to the real issue — which is how government regulators and trade negotiators write their rules.

That is what the sound-science bill, which was introduced by Stephen Fincher (Republican, Tennessee) and ended up in the House-passed version of the farm bill, tried to do. It would have covered every federal agency, although it originated in farmers' concerns about possible regulation of antibiotic use in animals. It required, among other things, that US regulators favour data from research that is "experimental, empirical, quantifiable, and reproducible". Which sounds fair enough, until you think about it: the approach would discount, for example, the use of weather modelling, or of data collected from one-off events, such as natural disasters.

Fortunately, scientific and environmental groups rumbled what the bill was really about, and got it turfed out of the farm bill just before it reached Obama's desk. However, dealing with such provisions is a bit like whack-a-mole. There is another mole already in sight on Capitol Hill: the Secret Science Reform Act, now under consideration by the House science committee, to stop the Environmental Protection Agency from using data that are not publicly available in its assessments.

And who could argue with that? Well, one issue with making all such data public is that it gives industry grounds for refusing to hand confidential data over, as it would then become public.

In the end, regulatory arguments are more philosophical than scientific in their nature. Environmentalists advocate caution in the face of uncertainty; industry wants cost-benefit analysis.

The natural sciences have little to say on which approach is wiser. Industry, however, has become adroit at using the concept of sound science to advocate the latter path. Too many researchers, as well as the wider public, are taken in by the claim that when someone says they are seeking the scientific answer to a regulatory question, they mean what they say. They very rarely do. ■

Colin Macilwain *writes about science policy from Edinburgh, UK.*
e-mail: cfmworldview@gmail.com

© 2014 Macmillan Publishers Limited. All rights reserved

RESEARCH HIGHLIGHTS

Selections from the scientific literature

PHOTONICS

Light goes one way on a chip

A device that controls light so that it travels in just one direction could be used in high-speed computers that carry signals using light, rather than electric charges.

A team led by Lan Yang and Şahin Kaya Özdemir at Washington University in St. Louis created a diode using two doughnut-shaped rings on a silicon chip. While one ring absorbs an incoming light signal, the other amplifies it. When the rings are close together, light travels through the device both ways. When the rings are farther apart, the signal can be amplified in one direction but blocked in the other.

The device is smaller and uses less power than existing optical diodes.

Nature Phys. **http://doi.org/r8n (2014)**



CRAIG MCCLAIN

ECOLOGY

Fallen trees form a sea-floor feast

Dead trees at the bottom of the ocean host a diverse range of bacteria, fungi and molluscs (pictured; a cent is included for scale).

Craig McClain of the National Evolutionary Synthesis Center in Durham, North Carolina, and James Barry of the Monterey Bay Aquarium Research Institute in Moss Landing, California, left 18 *Acacia* hardwood logs at a depth of more than 3,000 metres in the northeast Pacific Ocean and retrieved them after five years.

The duo found thriving ecosystems that varied dramatically between logs, even though the wood was within an area 500 metres square: on average, the logs were only about 25% similar in terms of species composition. Key colonizers were wood-boring bivalves, which create holes for other organisms to shelter in and provide food in the form of wood chips and faeces.

Changing patterns of deforestation, river flow and hurricanes might affect the frequency and size of such 'wood falls', which could have a significant impact on deep-sea diversity, the authors say.

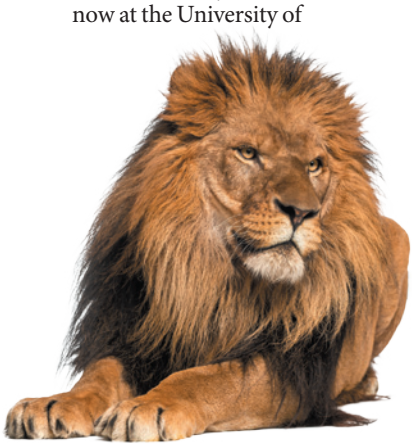
Biol. Lett. 10, **20140129 (2014)**

EVOLUTION

Ancient lion DNA yields family tree

Five genetically distinct lion populations roam in Africa and Asia — a finding that hints at greater diversity in these animals than previously thought.

A team led by Ross Barnett, now at the University of Copenhagen, analysed mitochondrial DNA (mtDNA) from the remains of 14 lions (*Panthera leo*) in museums, including extinct individuals from North Africa and Iran. Comparisons with mtDNA from other ancient and modern lions (**pictured**) suggested that the different populations are descended from an ancestral one living in southeastern Africa around 124,000 years ago. Habitat changes led to the expansion of the lions' range throughout Africa and, beginning around 21,000 years ago, into Asia and the Middle East.



Lion conservation and restoration efforts — which currently recognize only two populations in Africa and Asia — ought to account for this extra diversity, the authors say.

BMC Evol. Biol. 14, **70 (2014)**

CONSERVATION BIOLOGY

Unique birds top conservation list

An analysis of evolutionary relationships between all of the world's known birds prioritizes some of them for conservation on the basis of their genetic uniqueness.

Walter Jetz of Yale University in New Haven, Connecticut, Arne Mooers of Simon Fraser University in Burnaby, Canada, and their colleagues examined the family tree of almost 10,000 bird species. They calculated the birds' evolutionary distinctness — a measure of a species' separation from others on the family tree. Birds that ranked high in distinctness and that occupy small ranges or are threatened with extinction include the giant ibis (*Thaumatibis gigantea*) and the kakapo, a flightless parrot (*Strigops habroptilus*).

Many of these birds live outside well-known

LIFE ON WHITE/ALAMY

© 2014 Macmillan Publishers Limited. All rights reserved

biodiversity hotspots, suggesting that current conservation planning does not adequately cover these evolutionary rarities.
Curr. Biol. **http://doi.org/r9f (2014)**

CIVIL ENGINEERING

Seismic 'shield' stops quake shake

An array of deep holes in the ground seems to lessen shaking in certain locations during a simulated earthquake.

Stéphane Brûlé of soil-engineering company Ménard in Nozay, France, and his colleagues drilled a grid of boreholes 5 metres deep into the soil near Grenoble. A crane then lowered a probe into the ground nearby, where it vibrated to simulate an earthquake. The boreholes were located in places where the seismic waves interfered with one another and cancelled each other out. Thus, the array of holes deflected much of the seismic energy and bounced it back towards the source.

The approach could lead to new ways to protect buildings from shaking during earthquakes, the authors say.
Phys. Rev. Lett. 112, **133901 (2014)**

NEUROBIOLOGY

Surprising effects of prion drug

A chemical that combats pathogenic prion proteins in infected mouse cells worsens the problem in cells from other species. The finding could explain why the drug, quinacrine, has been ineffective in many clinical trials.

Prion infections turn healthy proteins into abnormally folded forms, which cause fatal neurodegenerative diseases in humans and other animals. Quinacrine reduces this misfolding in mouse cells, but Glenn Telling and his colleagues at Colorado State University in Fort Collins discovered that the drug has the opposite effect in deer, elk and moose cells, causing the proteins to misfold even more. The researchers think that small differences in prion protein sequences between species could be to blame.

Drug developers will need to rethink how they screen molecules for antiprion activity, the authors say.
Proc. Natl Acad. Sci. USA **http://doi.org/r9d (2014)**

MATERIALS

Printer squirts out nanotubes

Inkjet printers can produce thin films of carbon nanotubes for use as electrodes in stretchy electronic circuits.

Yongtaek Hong and his colleagues at Seoul National University printed layers of single-walled carbon nanotubes onto a stretchable silicon-based material. The authors found that the electrical properties of the films improved after washing them with water and soaking them in diluted nitric acid. Furthermore, a film with five layers of nanotubes performed better than a single layer, maintaining conductive properties under 100% tensile strain.

This method is simpler and more scalable than previous ones, the authors say.
Appl. Phys. Lett. 104, **113103 (2014)**

NEUROSCIENCE

Turn on the light to make myelin

Brain circuits change throughout life, and researchers in California have discovered a mechanism for one such change: the thickening of the myelin sheath that surrounds nerve fibres and helps neurons to fire.

Michelle Monje and her co-workers at Stanford University School of Medicine studied mice that had been engineered so that light can stimulate neurons in the brain's premotor cortex. This stimulation resulted in the generation of more cells called oligodendrocytes, which make myelin, and a thickening of the myelin sheath in this and other brain regions. Furthermore, mice that were stimulated with light and had thicker myelin showed better motor function than normal animals.

The results could point to ways of boosting myelin formation in neurodegenerative diseases such as multiple sclerosis, the authors say.
Science **http://doi.org/r9g (2014)**

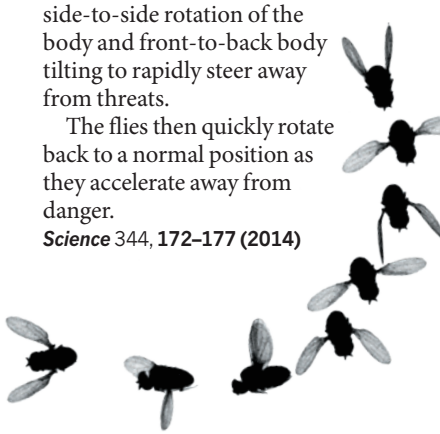
BIOMECHANICS

Fast moves of fleeing fruit flies

To dodge predators, flies in flight execute banked turns in just a few wingbeats — much faster than the steering motions that have been previously observed in flies.

Michael Dickinson and his colleagues at the University of Washington in Seattle used three high-speed cameras operating at 7,500 frames per second to capture the flight responses of a fruit fly species (*Drosophila hydei*; **pictured**) when it evades a threat looming in front of it. They found that with slight changes in wing motion, the insects use a combination of side-to-side rotation of the body and front-to-back body tilting to rapidly steer away from threats.

The flies then quickly rotate back to a normal position as they accelerate away from danger.
Science 344, **172–177 (2014)**



FLORIAN MUJURES

SOCIAL SELECTION

Popular articles on social media

Lionfish prove a main draw on Twitter

Lionfish took a bite of Twitter attention at the start of this month in a paper that reveals some unintended effects of population-control measures that could undermine future efforts.

Isabelle Côté at Simon Fraser University in Burnaby, Canada, and her team, looked at what happens after invasive predatory lionfish (*Pterois volitans/miles*) were culled at coral-reef sites in the Bahamas. The effects on surviving fish were marked: they became less active and hid deeper in the reefs during the day compared with fish in reefs where no culls took place. The wary lionfish also seemed able to spot approaching divers from a greater distance.

As Côté pointed out on Twitter, these changes in behaviour could lead to underestimating the size of the lionfish population.
PLoS ONE 9, **e94248 (2014)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/ccpzgi

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2014 Macmillan Publishers Limited. All rights reserved

SEVEN DAYS *The news in brief*

RESEARCH

Exotic particle

Data from the Large Hadron Collider beauty experiment (LHCb) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, have confirmed the existence of an exotic particle, Z(4430). The LHCb published the finding on 9 April. The particle is exotic because it may consist of four quarks — fundamental particles that are usually bound in twos or threes and are found in everyday matter. A number of experiments have observed such exotic particles over the past 11 years, but Z(4430) is only the second sighting to be confirmed.

Scientist apologizes

Japanese stem-cell scientist Haruko Obokata apologized on 9 April for errors in her work after being found guilty of research misconduct by her institute, the RIKEN Center for Developmental Biology, earlier this month. She took full responsibility for the errors — which included the use of a duplicated image — blaming her immaturity and ignorance of research protocols. She maintains that the research on turning mature cells into stem cells, published in *Nature*, is still valid and that the inaccuracies were not deliberate. See page 299 and go.nature.com/jblotc for more.

BUSINESS

Drug firms fined

On 7 April, a court fined two drug firms US$9 billion for not revealing links between a diabetes drug called Actos (pioglitazone) and bladder cancer. The Takeda Pharmaceutical Company, of Osaka, Japan, was ordered to pay $6 billion and Eli Lilly, of Indianapolis, Indiana, $3 billion. They were also ordered to pay $1.475 million in compensation. Takeda and Eli Lilly disagree with the verdict.



MIKE JANES/FOUR SEAM IMAGES VIA AP IMAGES

The dark glide of the Moon

A total lunar eclipse on 15 April was the first in a 'tetrad' — a set of four consecutive total lunar eclipses. Many in the Americas glimpsed the distinctive 'blood Moon', in which the Sun's visible red light is scattered and reflected from the Moon, which is otherwise completely masked in Earth's shadow (pictured, from Parrish, Florida). The remaining three eclipses in the tetrad will take place on 8 October, 4 April 2015 and 28 September 2015. Lunar eclipses occur around twice a year, but total lunar eclipses are less frequent, and only eight tetrads will occur in the twenty-first century; there were no tetrads from 1600 to 1990.

FACILITIES

Telescope shortlist

A committee announced the shortlist for one of two locations of the planned Cherenkov Telescope Array γ-ray detector on 10 April at a meeting in Munich, Germany. Representatives of the 12 governments that will fund the array selected Aar in southern Namibia and Armazones in Chile from a list of five for the larger site in the Southern Hemisphere. The final selection will take place by the end of the year. The funders have yet to whittle down a shortlist for the array's smaller site in the Northern Hemisphere. See page 297 and go.nature.com/n8posd for more.

POLICY

Immigration barrier

Fewer international students are coming to the United Kingdom to study science, technology, engineering and mathematics because of tough immigration rules, and prospective students are made to feel unwelcome in the country. So finds a report from the House of Lords Science and Technology Committee published on 11 April. The report says that student numbers have fallen by more than 10% over the past two years. The government maintained that there was no clear evidence that immigration rules were deterring international students (see *Nature* **506,** 14–15; 2014).

Medical history

A trove of biomedical journal archives will be digitized and made freely available by the US National Library of Medicine in Bethesda, Maryland, and the Wellcome Library in London. The project, announced on 14 April, is expected to take three years to complete. The collection will include publications that span the past 150 years, such as the *Transactions of the Epidemiological Society of London*, the *Indian Medical Gazette* and *Mental Hygiene*. The Wellcome

© 2014 Macmillan Publishers Limited. All rights reserved

Library will donate £750,000 (US$1.25 million) to the National Library of Medicine to support the project.

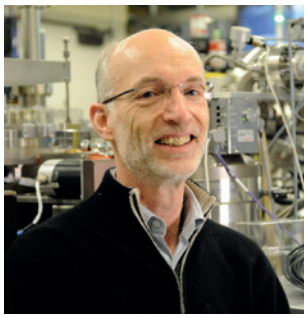
Nuclear restart

Three years after the Fukushima-Daiichi nuclear accident, the Japanese government has reapproved the use of nuclear power. An energy policy announced on 11 April says that nuclear reactors that meet tough regulatory standards will be reactivated. The plan reverses a decision by the previous government to phase out nuclear power. Japan's 48 nuclear reactors were shut down in the wake of the Fukushima disaster in March 2011.

Hepatitis C report

The World Health Organization unveiled its first global treatment guidelines for hepatitis C virus (HCV) on 9 April. The recommendations will guide health officials in treating the world's estimated 185 million cases of HCV. Newly available drugs can cure most infections, but the price of the drugs must be reduced, the report says. The virus is blood-borne, and is often transmitted through sharing injection needles and by poor sterilization of medical equipment. It causes cirrhosis and liver cancer. See page 295 for more.

AWARDS



TAF/RICK ENGLISH

Technology prize

British physicist Stuart Parkin (**pictured**) was awarded the 2014 Millennium Technology Prize on 9 April. The award recognizes Parkin, who is director of the Max Planck Institute of Microstructure Physics in Halle, Germany, for his work in spin electronics. His research has enabled a 1,000-fold increase in the storage capacity of hard-disk drives, and led to innovations such as cloud computing and online video streaming. The €1-million (US$1.4-million) prize is awarded every two years by the Technology Academy Finland.

PEOPLE

Health chief quits

The US secretary of the Department of Health and Human Services, Kathleen Sebelius, resigned on 11 April after the troubled roll-out of sweeping health-care reforms. US President Barack Obama nominated Sylvia Burwell, director of the Office of Management and Budget, as her replacement. Sebelius notably overruled the scientific advice of the US Food and Drug Administration in 2011 when she sought to restrict access to the 'morning after' pill (levonorgestrel) for girls aged under 17, although a judge later ordered that it must be available over the counter for all ages (see *Nature* **496,** 138; 2013).

Plagiarism battle

Germany's former education and research minister, Annette Schavan, announced on 10 April that she will give up the fight to have her doctorate reinstated after the University of Düsseldorf withdrew it on the grounds of plagiarism. The university revoked her title last year after identifying parts of her PhD thesis that had been copied from sources that had not been correctly cited. On 20 March, a court rejected her claim that the university handled her case unfairly. She will not appeal the court ruling.

Stem-cell move

The former stem-cell chief at the US National Institutes of Health (NIH) will join the New York Stem Cell Foundation, a non-profit organization that funds research into the technology, it was announced on 9 April. Mahendra Rao resigned last month as director of the NIH Center for Regenerative Medicine, saying that the NIH was not pushing through enough stem-cell therapies to clinical trial. See go.nature.com/eptx5e for more.

China investigation

Shen Weichen, a high-ranking politician at China's national science association, is under investigation for "suspected serious violation of discipline and laws", Chinese state media reported on 12 April. Shen has been Communist Party secretary at the China Association for Science and Technology in Beijing since April 2013. He is also a member of the organization undertaking the probe: the party's Central Commission for Discipline Inspection, which has been leading the recent crackdown on corruption, state media said.

COMING UP

28 APRIL–7 MAY
Countries that are signatories to the Antarctic Treaty meet in Brasília to discuss how the treaty is being implemented. Talks at the meeting will including how to manage access for research, environmental protection and tourism.
go.nature.com/e7jfwq

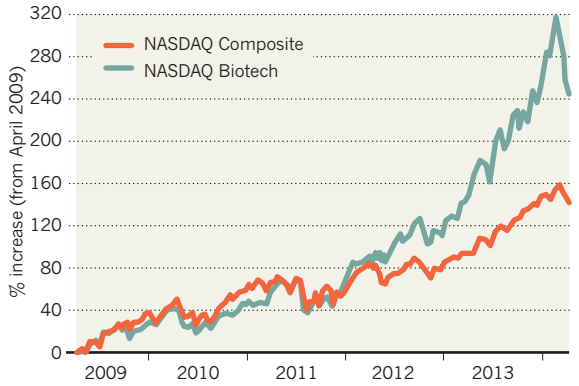
23–25 APRIL
Researchers at the Stem Cell Summit 2014 in Cambridge, Massachusetts, will discuss developments in stem-cell therapies, regenerative medicine and regulation of stem-cell research.
go.nature.com/ui4q3m

TREND WATCH

After a promising start to 2014, biotechnology stocks are taking a tumble. The NASDAQ Biotechnology Index (NBI), an amalgam of 121 biotechnology and pharmaceutical companies, has fallen by 21% since its peak on 25 February. The broader NASDAQ Composite Index fell by 6% over the same period. The NBI fall has befuddled analysts, and spurred fears that a biotechnology bubble is bursting (see *Nature* **500,** 513; 2013). But it may simply be a correction to an especially sharp rise that began in January.

BIOTECH BUBBLE BURSTING?

An index fund of biotechnology companies had soared above the NASDAQ Composite Index, but is now falling back.



SOURCE: NASDAQ

NATURE.COM
For daily news updates see:
www.nature.com/news

© 2014 Macmillan Publishers Limited. All rights reserved

NEWS IN FOCUS

GENETICS How trauma leaves its mark through the generations — via sperm **p.296**

STEM CELLS Tests proceed on hotly debated method as author stands firm **p.299**

SPACE Cut-price, mini satellites swarm into orbit and maybe beyond **p.300**

ENERGY Companies sink millions into quest for marine power **p.302**

SAURABH DAS/AP



Campaigners in India have been calling for access to cheap hepatitis C treatments.

PUBLIC HEALTH

Hepatitis C drugs not reaching poor

Treatment guidelines for virus highlight challenge of paying for expensive drugs in low-income countries.

BY EWEN CALLAWAY

The publication last week of the first treatment guidelines for hepatitis C virus (HCV), and the advent of drugs that can cure most infections of the virus, have left public-health researchers with a touch of déjà vu.

Three decades after wrestling to lower the cost of AIDS drugs (prices fell from about US$10,000 per patient per year in the 1990s to less than $100 in the mid-2000s), they are once again asking how expensive life-saving medicines can be made affordable for patients.

"We've been here before," says Stefan Wiktor, head of the global hepatitis programme at the World Health Organization (WHO) in Geneva, Switzerland, and lead author of the guidelines (see go.nature.com/bwhguw). And a similar success could follow, adds Andrew Hill, a pharmacologist at the University of Liverpool, UK, "if there's the political will to do it".

Discovered in 1989, HCV infects an estimated 185 million people worldwide, more than 80% of whom live in low- and middle-income countries (see 'Viral load'). Around 15% of Egypt's population, for example, is infected. The blood-borne virus is a leading cause of liver cancer and cirrhosis.

Most of these HCV cases had been ignored, mainly because available treatments did not work well and were relatively expensive. The combination of an antiviral drug, ribavirin, and an immune-system modulator called interferon cures less than 50% of patients after a year, while causing fatigue, nausea and even depression.

But that is now set to change. First, a new breed of HCV medicine that directly attacks the virus — approved by US and European regulators in 2011 — pushed cure rates to around 75%, although it did not fully eliminate the need for the ribavirin–interferon therapy. And in recent months, regulators have approved two drugs that can cure most HCV infections after three months and do not require interferon in many cases.

But the pricing of these new drugs has raised eyebrows even in Western countries, where private health insurers and public health-care systems tend to bear the costs of drugs. Sofosbuvir, made by Gilead Sciences in Foster City, California, debuted last December at $1,000 a pill — $84,000 for a 12-week course. And Janssen Pharmaceuticals, based in Beerse, Belgium, has priced its simeprevir at $66,000 for the same treatment period. Those prices are a non-starter for low- and middle-income countries. But it is not clear how the price will fall to a level that such countries can afford, says Wiktor.

Antiretroviral drugs that keep HIV in check reached less-wealthy countries only after pressure from non-governmental agencies, academics and patients. Generic-drug manufacturers showed that they could make the medicines cheaply, and funders such as the Global Fund to Fight AIDS, Tuberculosis and Malaria stepped in to pay for them. Some countries, such as Brazil and Thailand, even issued 'compulsory licences' to manufacture generic versions of HIV drugs, effectively ignoring patents.

These options will be open to HCV drugs, says Wiktor. "Competition and generic production really are the keys to reductions in prices."

In March, Gilead announced a deal with Egypt to provide sofosbuvir at $900 for a 12-week course. "I think it's an excellent first step, but it's only a first step," says Hill. He adds that treating every case of HCV in Egypt would far outstrip the country's annual health-care budget of around $4.5 billion.

Gregg Alton, executive vice-president of corporate and medical affairs at Gilead, says that the company is in negotiations with several Indian manufacturers to produce a generic version of sofosbuvir. But, according to a report from the Paris-based charity Médicins du Monde, the generic drug would be available to 60 mostly low-income countries, and not to many middle-income countries with high HCV burdens. Meanwhile, the Initiative for Medicines, ▶

© 2014 Macmillan Publishers Limited. All rights reserved

Access & Knowledge, a non-profit group in New York, has filed a lawsuit with India's patent office seeking to prevent Gilead from receiving a patent for sofosbuvir.

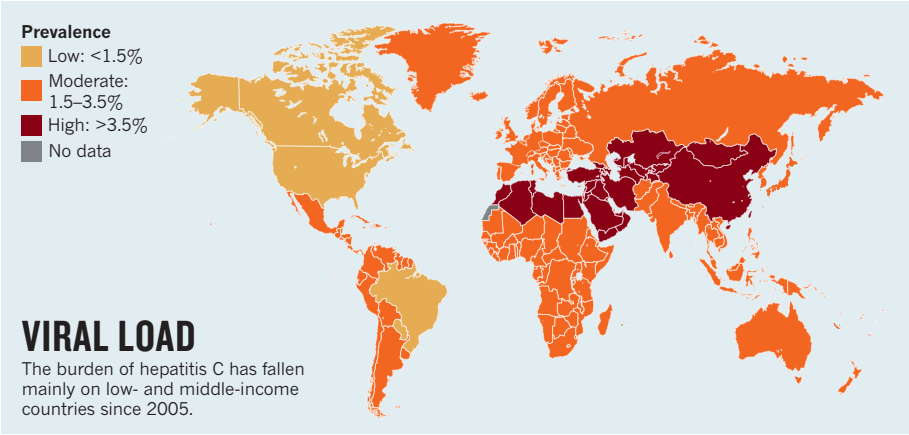
Cheap generic HCV drugs are within reach. In an analysis published in February in *Clinical Infectious Diseases*, Hill and his team compared the cost of producing generic HIV medicines and applied this analysis to the potential cost of HCV drugs (A. Hill *et al. Clin. Infect. Dis.* **58,** 928–936; 2014). They estimate that generic-drug manufacturers should be able to produce the pills at $100–250 for a 12-week course. But even at those prices generic manufacturers won't manufacture HCV medicines until there is large demand, Hill says.

Major donors have not yet signalled their intention to pay for the drugs. Geneva-based UNITAID welcomed the new WHO guidelines, but noted that access to drugs will depend on the development of cheap diagnostics. "It's possible it's really going to be up to the countries to fund their own programmes, which is really going to be a limiting factor," says Wiktor.

Jennifer Cohn, medical director of the drug-access programme for the charity Médecins Sans Frontières in Geneva, notes that middle-income countries face the highest hurdles. Increasingly, international donors are withdrawing support from these countries, and drug makers such as Gilead are likely to see them as untapped markets for their products and thus exclude them from generic licensing deals.

Cohn says that these countries should consider compulsory licensing, just as some countries did for HIV drugs. This practice is sanctioned by the World Trade Organization in certain cases — for example, to ensure affordable medicine prices — to protect public health.

But cost is not the only barrier. Most countries' health systems are not equipped to widely diagnose the disease or deliver the drugs. Wiktor says that it will be a missed opportunity if the global health community does not work out how to get the drugs to patients. "People are dying from liver cancer, from cirrhosis," he says, "and these deaths can be prevented." ■



VIRAL LOAD
The burden of hepatitis C has fallen mainly on low- and middle-income countries since 2005.

SOURCE: K. M. HANAFIAH *ET AL. HEPATOLOGY* **57,** 1333–1342 (2013)

EPIGENETICS

Sperm RNA carries marks of trauma

Stress alters the expression of small RNAs in male mice and leads to depressive behaviours in later generations.

BY VIRGINIA HUGHES

Trauma is insidious. It not only increases a person's risk for psychiatric disorders, but can also spill over into the next generation. People who were traumatized during the Khmer Rouge genocide in Cambodia tended to have children with depression and anxiety, for example, and children of Australian veterans of the Vietnam War have higher rates of suicide than the general population.

Trauma's impact comes partly from social factors, such as its influence on how parents interact with their children. But stress also leaves 'epigenetic marks' — chemical changes that affect how DNA is expressed without altering its sequence. A study published this week in *Nature Neuroscience* finds that stress in early life alters the production of small RNAs, called microRNAs, in the sperm of mice (K. Gapp *et al. Nature Neurosci.* http://dx.doi.org/10.1038/nn.3695; 2014). The mice show depressive behaviours that persist in their progeny, which also show glitches in metabolism.

The study is notable for showing that sperm responds to the environment, says Stephen Krawetz, a geneticist at Wayne State University School of Medicine in Detroit, Michigan, who studies microRNAs in human sperm. (He was not involved in the latest study.) "Dad is having a much larger role in the whole process, rather than just delivering his genome and being done with it," he says. He adds that this is one of a growing number of studies to show that subtle changes in sperm microRNAs "set the stage for a huge plethora of other effects".

In the new study, Isabelle Mansuy, a neuroscientist at the University of Zurich, Switzerland, and her colleagues periodically separated mother mice from their young pups and exposed the mothers to stressful situations — either by placing them in cold water or physically restraining them. These separations

MORE ONLINE

TOP NEWS



KATIE COLLINS/PRESS ASSOCIATION IMAGES

Report describes Tamiflu stockpiling as wasteful go.nature.com/m3glux

MORE NEWS

- Ancient Mars was too cold for persistent liquid water go.nature.com/sg3ijf
- Turkey eggs have antibacterial nanoparticles go.nature.com/soymor
- IceCube detector catches glimpse of another high-energy neutrino go.nature.com/h1sk8u

NATURE PODCAST



Sustainable energy for all; egg meets sperm; and island biodiversity nature.com/nature/podcast

© 2014 Macmillan Publishers Limited. All rights reserved

occurred every day but at erratic times, so that the mothers could not comfort their pups (termed the F1 generation) with extra cuddling before separation.

When raised this way, male offspring showed depressive behaviours and tended to underestimate risk, the study found. Their sperm also showed abnormally high expression of five microRNAs. One of these, miR-375, has been linked to stress and regulation of metabolism.

The F1 males' offspring, the F2 generation, showed similar depressive behaviours, as well as abnormal sugar metabolism. The F1 and F2 generations also had abnormal levels of the five microRNAs in their blood and in the hippocampus, a brain region involved in stress responses. Behavioural effects persisted in the F3 generation as well.

Mansuy and her team are now looking into whether similar microRNA biomarkers occur in people exposed to traumatic events — or in their children. "If some are altered persistently in blood, then they could be used as markers for susceptibility to stress or for developing psychiatric disorders," she says.

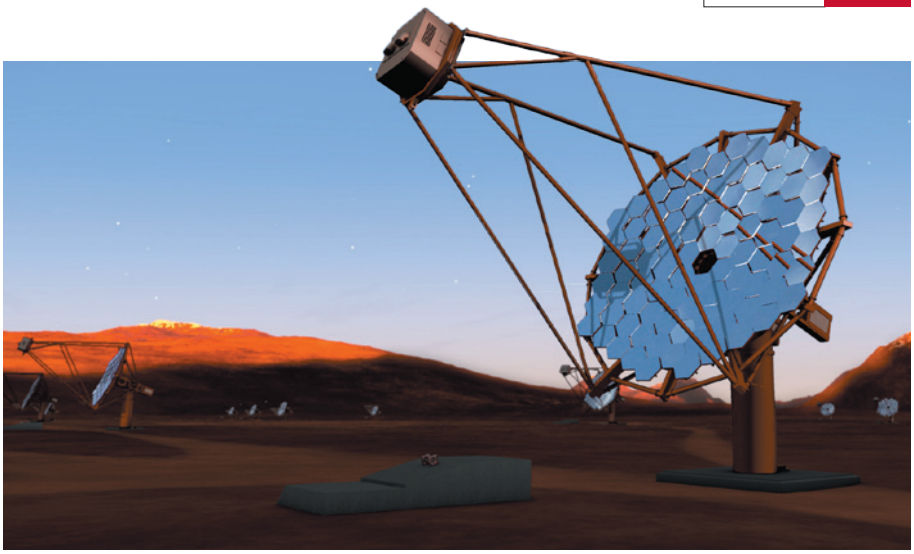
To rule out the possibility that the effects of stress were transmitted socially, the researchers also collected RNA from the F1 males' sperm and injected it into freshly fertilized eggs from untraumatized mice. This resulted in mice with comparable depressive behaviours and metabolic symptoms — and the depressive behaviours were passed, in turn, to the next generation.

The authors readily admit that there is much to be discovered about the biological underpinnings of these findings. No one knows how stress triggers the changes in sperm microRNA, for example. One potential route is through glucocorticoid receptors, proteins involved in the stress response that are expressed in sperm. It could be that stress hormones circulating in the blood make their way to the testes and bind to these receptors, somehow triggering changes in microRNA expression, says Sarah Kimmins, an epigeneticist at McGill University in Montreal, Canada. "Nobody's explored that, and I think that's a really exciting avenue."

"Dad is having a much larger role ... than just delivering his genome."

Krawetz notes another puzzle raised by the study — the stressful experience did not affect the sperm microRNA of the F2 or F3 generations. This could mean that the abnormalities in these progeny came from some other epigenetic mechanism, such as DNA methylation or chemical marks on histones, the proteins that DNA wraps around. But for now, he says, "that's all hand-waving". ■

DESY/MILDE SCIENCE COMM./EXOZET



The telescope array (artist's impression) will be split across the Northern and Southern hemispheres.

ASTRONOMY

Panel homes in on sites for γ-ray detector

Cherenkov Telescope Array will track high-energy photons to probe black holes, dark matter and relativity.

BY ELIZABETH GIBNEY

When very-high-energy γ-rays slam into Earth's atmosphere, they trigger particle showers that emit a faint blue light. With this light, astronomers want to trace the rare γ-rays — only a few strike each square metre of the atmosphere each month — back to their sources, violent objects such as supermassive black holes. But first researchers must find a home for the planned €200-million (US$277-million) Cherenkov Telescope Array (CTA) — or rather, two homes. The telescope will be made up of a 19-dish array in the Northern Hemisphere and a 99-dish array in the south.

At a meeting in Munich, Germany, on 10 April, representatives from the 12 CTA partner countries inched closer to picking the sites. In the Southern Hemisphere, they narrowed the list down to two possibilities: Aar, in southern Namibia; and Cerro Armazones in Chile's Atacama Desert. In the north, four sites remain in the running: two in the United States and one each in Mexico and Spain.

Some had hoped the panel would pick firm favourites. Last year, a committee of CTA scientists came up with a broader list of sites based on environmental factors such as weather and earthquake risk. The latest decision adds considerations such as political stability and the financial contributions of host nations. "The process is going slower than we'd like, but it's going, and that's great," says Rene Ong, a physicist at the University of California, Los Angeles, who has helped to plan for the CTA.

The array would study photons in an as-yet unexplored energy region: up to 100 teraelectronvolts. Cosmic rays — protons and other nuclei — emit these photons when they are accelerated at the surface of neutron stars and black holes, and when they collide in stellar winds.

The CTA would focus on the centre of the Milky Way because of the dark matter thought to lurk there; many theories predict that dark-matter particles could annihilate each other and emit γ-rays that the CTA should detect. The array would explore physics at energy scales well beyond the scope of most powerful accelerators.

The CTA would also probe theories of quantum gravity, which try to reconcile quantum mechanics with Einstein's theory of gravity. Some theories predict that very-high-energy photons, with wavelengths approaching the foamy quantum scale of space-time, will travel slightly slower than lower-energy photons from the same source. Observations of γ-rays at different energies could reveal arrival-time lags.

The CTA panel aims to pick a final southern site by the end of the year. Choosing the northern site may take longer, says panel chair Beatrix Vierkorn-Rudolph, deputy director-general of Germany's Federal Ministry of Education and Research. Astronomers hope to be ready to start construction by the end of 2015 and to begin full operations in around 2020. ■

© 2014 Macmillan Publishers Limited. All rights reserved

CLIMATE POLICY

IPCC report under fire

Critics attack panel's lack of specific guidance on how countries should lower emissions.

BY QUIRIN SCHIERMEIER

Is the Intergovernmental Panel on Climate Change (IPCC) losing steam? It is tackling a global problem that remains as urgent as ever. But the expert body's latest report, released on 13 April and focused on mitigation, has left many climate experts and policy analysts unsatisfied.

The document, a policy summary prepared by Working Group III of the IPCC, is the third instalment of the IPCC's Fifth Assessment Report. It follows reports on the science of climate and the impacts of climate change, released in the past few months. Compiled by hundreds of lead and reviewing authors over several years, the report warns that without substantial policy and technology changes, the world is heading towards dangerous temperature rises. Its focus is therefore on the technological and economic options for stabilizing atmospheric greenhouse-gas concentrations at acceptable levels.

"The analysis is all there," says Bob Ward, a policy director at the London School of Economics' Grantham Research Institute on Climate Change and the Environment, who was not involved in the report. "Now leaders must get to grips with that message."

But others find the key conclusions unsurprising and short of detail. They say that the document sidesteps any hint of what specific countries, or groups of countries, should do to move towards clean energy systems. In particular, critics are concerned about the large emitters — China, the United States, the European Union and India, which together account for more than half of global carbon dioxide output.

"The core message is that emissions need to go down, and that the costs are affordable," says Glen Peters, a climate-policy analyst at the Center for International Climate and Environmental Research — Oslo (CICERO). "But the big-picture stuff is not very helpful for decision-makers in specific countries, and it is pretty much useless for the international climate-negotiation process."



PAUL MCGEE/MOMENT/GETTY

Renewable-energy systems are improving globally.

Others agree. Steve Rayner, who studies policy frameworks for climate change at the University of Oxford, UK, says: "The headline findings are predictable — and not very exciting. The all-too-familiar-sounding bottom-line messages don't seem to justify the huge effort involved."

The report, released at a packed press conference in a Berlin hotel, details how annual emissions have increased from 27 gigatonnes (Gt) of CO_2 equivalent in 1970 to 49 Gt in 2010 (see 'Who emits what?'). Emissions may now have reached more than 52 Gt per year. The report adds that to have a 50% chance of keeping the global surface-temperature increase below the UN target of 2 °C, humans must not release more than an additional 1,550 Gt of greenhouse gases before 2100. At current rates, that limit will be exceeded before 2050.

The working group also warns that developed countries' emissions targets for 2020 — agreed at a 2010 climate summit in Cancun, Mexico — are inconsistent with the 2 °C ceiling, which was set at the same summit. "Meeting this goal would require further substantial reductions beyond 2020," the report finds.

The IPCC also acknowledges that the renewable-energy industry is making substantial gains in performance and cost reductions. But it notes that growing global energy demand and an increase in coal's share of the global fuel mix in recent years threaten to thwart mitigation efforts.

"Irrespective of precise emissions targets, we have to start to bring the mitigation train on track by fundamentally up-scaling low-carbon forms of energy around the world," says Ottmar Edenhofer, co-chair of the IPCC's mitigation group. Nuclear power and environmentally safe geoengineering technologies, such as the removal of atmospheric CO_2 and the capture and storage of carbon emissions, will all need to be considered, he says.

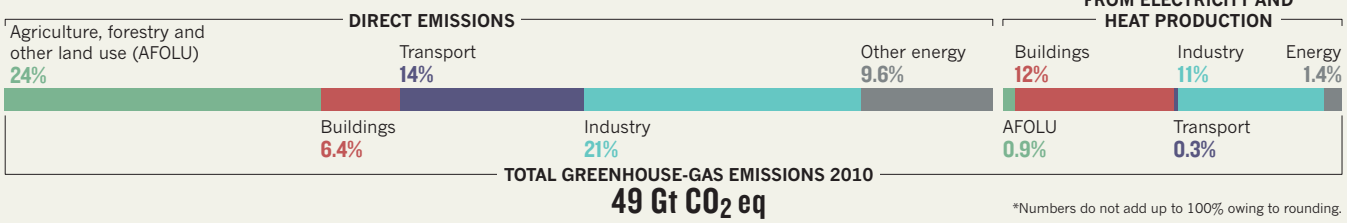
Some researchers have long argued for a more pragmatic and diversified approach to climate change. For example, one group wrote in a policy paper in 2010 that fostering technological progress while focusing on poverty reduction — an estimated 1.5 billion people have no access to electricity — might ultimately prove more effective than international treaties such as the expired Kyoto Protocol on climate change (G. Prins *et al*. *The Hartwell Paper*; LSE, 2010).

Robert Stavins, an environmental economist at Harvard University in Cambridge, Massachusetts, and a lead author of the IPCC report, emphasizes that the greatest challenges in mitigating the effects of climate change will be political, not technological.

Most critics agree that the IPCC, despite not having a remit to prescribe policies, has managed to strengthen the links between science and politics. "The process forces policy-makers to really engage with the science underlying climate change," says Peters. "Given the scale of the problem we're facing, that's good to know." ■

WHO EMITS WHAT?

Anthropogenic greenhouse-gas emissions reached 49 gigatonnes of carbon dioxide equivalent in 2010. Here, the total is broken down by economic sector and electricity and heat production*.



*Numbers do not add up to 100% owing to rounding.

SOURCE: IPCC

© 2014 Macmillan Publishers Limited. All rights reserved

KIMIMASA MAYAMA/EPA/ALAMY



At a press conference last week, Haruko Obokata insisted that her contentious technique does work.

REPRODUCIBILITY

Biologist defiant over stem-cell method

Japanese author of controversial papers denies wrongdoing and stands by results as testing of her protocol begins.

BY DAVID CYRANOSKI

The lead author of two hotly debated stem-cell papers made a tearful plea for forgiveness last week after her employer found her guilty of misconduct. Haruko Obokata, a researcher at the RIKEN Center for Developmental Biology (CDB) in Kobe, Japan, struggled to answer questions about errors in the papers, which described how simple stressors such as acid or pressure could reprogram mature cells into an embryonic-like state. But that did not stop her from insisting that the reports were not fraudulent and that the phenomenon described in them is real.

Her comments have left observers wondering about the outcome of a controversy that has raged since the papers were published in *Nature* in January[1,2]. Clarity on the claimed creation of STAP cells (for stimulus-triggered acquisition of pluripotency) now awaits three key events, all expected in the next few months. Stem-cell scientists hope that one of these — a replication attempt based on Obokata's protocol, by Hitoshi Niwa, a co-author of the papers who also works at the CDB — will be conclusive.

"This looks like a rigorous protocol that hopefully will settle the question of whether pluripotent STAP cells can be generated or not," says Rudolf Jaenisch, a stem-cell biologist at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, who was shown Niwa's protocol by *Nature*. (*Nature*'s news and comment team is editorially independent of its research editorial team.)

Problems with the papers, including accusations that Obokata had plagiarized passages of text and used duplicated images, arose soon after publication. Moreover, other groups said that they were unable to reproduce the results.

RIKEN decided to investigate, and on 1 April reported a number of uncomfortable findings (see *Nature* http://doi.org/sbb; 2014). Two problems were deemed misconduct: the re-use of an image that Obokata had included in her 2011 doctoral dissertation to describe different kinds of cells from those described in the STAP papers, and an image of an electrophoresis gel that had been spliced into another image, making it appear to be part of a different experiment.

Obokata fought back. In a statement on 1 April, she accused the RIKEN committee of giving her no chance to explain how those mistakes were made. Then, on 8 April, she submitted an appeal asking RIKEN to withdraw the charges and assemble another committee to investigate. At a press conference she held on 9 April, she passionately made her case and stated that she had succeeded in creating STAP cells more than 200 times. She blamed the misconduct findings on personal failings. "My immaturity and lack of training, it's really shameful," she said. "But with my lawyer's help, I do think I'll be able to dispel these suspicions."

RIKEN has 50 days to determine whether its original misconduct findings should be upheld. But, either way, the committee has made it clear that its conclusions do not address the matter of whether the experimental results were valid.

Weighty evidence on that point could come as early as this summer. Teruhiko Wakayama at the University of Yamanashi, Japan, a co-author on the papers, has sent eight supposed STAP stem-cell lines to an undisclosed genetic-testing centre to determine whether they could have been produced by contamination rather than stress. Because several mice were used in the work, one would expect to see genetic differences between the STAP lines — and for those differences to correspond to the animals from which the cells came. But if all the lines are genetically identical, it would suggest that the original results came about through contamination.

Evidence on whether STAP works at all may emerge even earlier. On 1 April, Niwa began tests approved by RIKEN that will follow the STAP method. He will use the same mouse cell types (lymphocytes) that Obokata did, as well as other kinds of mature mouse cells, and will perform some additional tests to try to trace the conversion. The work is expected to take a year, but an interim report will be released around the end of July. This could show whether the cells have passed the first crucial stages of STAP conversion. RIKEN says that Obokata will not be a member of the re-evaluation team, although she might be consulted on technical details.

Some researchers are hopeful that Niwa's study will clear up the controversy. "Hitoshi Niwa's reputation is stellar," says Hans Schöler of the Max Planck Institute for Molecular Biomedicine in Münster, Germany. He suggests that Niwa should involve Obokata, "following her step by step and letting her make all the solutions with his reagents in his presence, to see if she forgot to mention a critical step".

Josh Brickman, of the Danish Stem Cell Center at the University of Copenhagen, says that although negative results should settle the issue, "a positive replication, or worse, a limited positive replication" would just lead to more questions. "I would not be surprised if stress-induced reprogramming ends up being partially reproducible — an extremely rare and fickle process," he says. "If this is the case, it may become increasingly difficult to determine exactly what was right in the original papers." ■

1. Obokata, H. *et al. Nature* **505,** 641–647 (2014).
2. Obokata, H. *et al. Nature* **505,** 676–680 (2014).

© 2014 Macmillan Publishers Limited. All rights reserved



NASA

Two 30-centimetre-long Earth-observing CubeSats were released from the International Space Station in February.

SPACE

Mini satellites prove their scientific power

Proliferation of 'CubeSats' offers fresh and fast way to gather space data.

BY NICOLA JONES

A tiny box orbiting Earth is sending home big data. Built mostly by undergraduate students for less than US$1.5 million, the 10 centimetre × 10 cm × 30 cm satellite, called Firefly, is now beaming back information on terrestrial γ-ray flashes — energetic bursts that are triggered by lightning and fired out into space.

Firefly, which launched in November, is the first satellite devoted to studying the flashes, and scientists are eagerly awaiting the results. "I think this will be really amazing," says Therese Moretto Jorgensen, a programme director in the geospace sciences division of the National Science Foundation (NSF) in Arlington, Virginia.

Firefly is part of a growing band of mini satellites known as CubeSats that are now coming into their scientific own. Thanks to cheap parts and free launches, CubeSat launches are booming. As scientists and engineers prepare to meet next week for their annual CubeSat Developers' Workshop in San Luis Obispo, California, they are celebrating a record 63 deployments last year, nearly half the total of about 130 since launches began in 2003 (see 'Good things in small packages'). Experts say that the machines are at last starting to tackle valuable science questions. "After ten years, they're finally doing something more than just saying, 'Hi, I'm alive'," says space historian Jonathan McDowell at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

CubeSats have been deployed for everything from detecting the electromagnetic signals that might accompany earthquakes to monitoring the production of bacterial proteins in space. One mission is even planning to be the first to leave Earth's orbit: INSPIRE, a pair of CubeSats that could launch as early as the end of this year. It will trek 1.5 million kilometres — 1% of the distance to the Sun — and, from there, map the charged particles of the solar wind that shoot out from the Sun.

CubeSats were invented in 1999 as an educational tool. They were designed to be small and simple enough for graduate students to build, and they had a standard size, to make it easy for rockets to accommodate them. Robert Twiggs, an aerospace engineer at Stanford University in California, and the satellite department at the California Polytechnic State Institute in

© 2014 Macmillan Publishers Limited. All rights reserved

San Luis Obispo designed a 10-cm-a-side box: just big enough for a communications unit, solar panels, a battery and a few extras. The total package weighs in at about 1 kilogram. The first CubeSats launched in 2003, when six projects hitched a ride on a Russian Eurockot launch vehicle for about $40,000 per cube.

The format took off. In 2010, NASA started a CubeSat Launch Initiative, which each year offers a free ride to a dozen or more successful applicants as hitchhikers on primary loads. The initiative has awarded launch spots to more than 100 projects, although only 24 have been successfully shot into orbit so far (the rest are on a waiting list).

INCREASING DEMAND

In 2012, another avenue into space opened up: deployment from the International Space Station (ISS). And CubeSat builders can continue to pay commercial launch suppliers such as SpaceX for rides that typically cost $80,000 per cube. The price has gone up since 2003, says Twiggs, because of increasing demand.

Meanwhile, about a dozen shops have sprung up to sell standardized parts. At one, Pumpkin in San Francisco, California, a chassis costs about $7,500; a solar panel around $2,500. An entire CubeSat can be built for less than $100,000, says Twiggs. "It depends whether you have free student slave labour or not."

Such low costs prompted the NSF to start a programme in 2007 to encourage the use of CubeSats as a way to get more data more quickly than multimillion-dollar satellite missions. "It started, really, out of curiosity," says Jorgensen, who manages the NSF CubeSat programme and who has been pleasantly surprised by the results. "In terms of the data we want for space weather, there's actually very little that we cannot fit on a CubeSat." The programme has funded ten missions so far, including Firefly, at about $900,000 each.

The NSF's showcase mission was the RAX-2 satellite, which, like Firefly, was a three-CubeSat unit. It launched in 2011 and for 18 months measured ionosphere turbulence caused by solar storms. RAX-2 was designed to measure scattering in the signals from powerful ground-based radar transmitters that pass into space through the ionosphere.

The mission produced surprising data about how electrons in the upper atmosphere heat up, which should improve models of the effects that space weather has on satellites, says Jorgensen (H. Bahcivan *et al. J. Geophys. Res.* **119,** 1233–1249; 2014). The RAX team is now putting together a proposal for a follow-on CubeSat mission, which would monitor the ionosphere's scattering of television signals.

The INSPIRE project, more expensive at $5.5 million, will monitor solar effects closer to the Sun. As the satellite pair slowly separates, the units will map the turbulence in the solar wind at different scales. It will be the first time that turbulent features smaller than 10 kilometres have been measured, says Andrew Klesh, principal investigator of the INSPIRE project at the Jet Propulsion Laboratory in Pasadena, California. But much of the mission's value, he says, will lie in simply proving that a CubeSat can survive, navigate and communicate outside Earth's orbit. To ready the satellites, the team had to shrink a long-distance radio by an order of magnitude and build tiny cold-gas-firing thrusters, all in less than 20 months.

Private companies are also investing in CubeSats. Planet Labs in San Francisco — founded in 2010 by three former NASA scientists — launched an Earth-imaging system in February. The constellation consists of a record 28 satellites, each a three-cube unit, released from a dock on the ISS. By using so many satellites, says Twiggs, scientists can in theory get data on a given spot on Earth more than once a day, radically improving real-time monitoring. "This is the killer application for CubeSats," he says. But without thrusters, it is unclear how well these satellites will spread out before they fall back to Earth, probably within a year. The company declined to comment on its data so far.

"I love the idea that I can basically get something at Walmart and fly it in a satellite."

SMALLER STILL

Leaders in the field are now pushing for a smaller standard size. In 2009, Twiggs slimmed a CubeSat down into a 5-cm-a-side 'PocketQube'; the first four launched in November 2013. One of these demonstration projects, jokingly named $50SAT, cost just a few hundred dollars in parts — including a $14 chip for a two-way radio. "I love the idea that I can basically get something at Walmart and fly it in a satellite," says Twiggs. For now, the smaller size lacks the larger model's economies of scale in terms of parts and launch opportunities.

Some are favouring even smaller devices. On 18 April, a SpaceX rocket is set to take a $30,000 Kickstarter-funded CubeSat called KickSat to the ISS. About two weeks later, after KickSat is ejected from the ISS, 104 postage-stamp-sized 'ChipSats' will be released. Fitted with radio transmitters, the ChipSats were offered up for auction for about $300 each to anyone who wanted one.

For now, chip- and pocket-sized satellites can do little more than broadcast their location or a short message. But "things will evolve from gimmick to return fairly quickly", predicts Jason Crusan, director of NASA's Advanced Exploration Systems Division in Washington DC. Swarms of such satellites could provide a kaleidoscopic view of a target, or be treated as disposable missions.

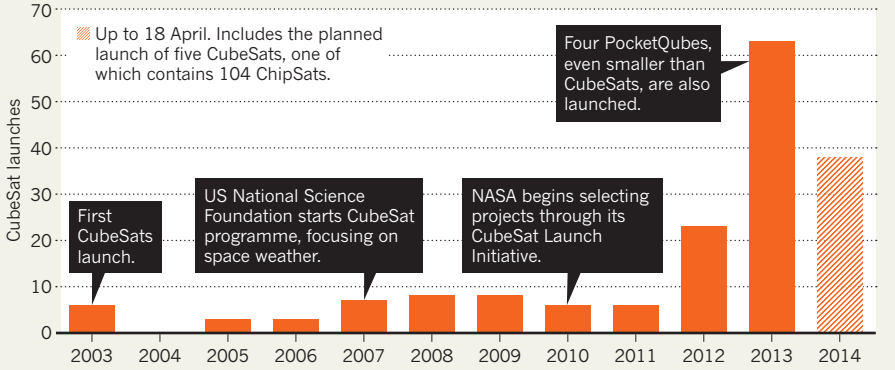
Michael Johnson, a physicist and visiting fellow at the University of Bristol, UK, says that one idea is to study the rings of Saturn. "You can do that with a billion-dollar satellite that has to stay far away to be safe," says Johnson, who is co-creator of the KickSat mission and founder of the company PocketSpacecraft.com. "Or you can fly 10,000 low-cost craft and land one" on one of the bigger pieces of rock in the rings, he says.

Johnson says that PocketSpacecraft.com is designing ChipSats so thin that they will not burn up on re-entry. That might allow researchers to map winds in Earth's upper atmosphere, or land a network of accelerometers on the surface of Mars.

Jorgensen is yet to be convinced of the ChipSats' usefulness. Still, she is keeping an open mind. "It's hard to imagine they will be capable enough," she says, "but that's exactly what people said about CubeSats." ■

GOOD THINGS IN SMALL PACKAGES

Launches of mini satellites called CubeSats reached a high last year, thanks to low-cost, standardized parts and increased deployment opportunities.



CORRECTION

The photo that accompanied the News story 'Ukraine's science in turmoil' (*Nature* **508,** 15–16; 2014) was not of the Crimean Astrophysical Observatory as stated but of a military radar station in the region. The correct image of the observatory can be seen at go.nature.com/elhdtf.

© 2014 Macmillan Publishers Limited. All rights reserved



Blue energy

After years in the doldrums, the quest to harvest energy from the oceans is gathering speed.

BY JEFF TOLLEFSON

Several kilometres off the coast near Perth in Western Australia, hidden beneath the waves and out of sight of watchful boaters, three giant buoys will soon begin producing electricity as they bob to the rhythm of the Indian Ocean. At 11 metres wide and 5 metres tall, the squat orange floats look a bit like giant pumpkins. As waves pass by, the tethered buoys will drive hydraulic pumps on the sea bottom, converting the motion of the ocean into 720 kilowatts of electricity to power a nearby naval base.

Carnegie Wave Energy of North Fremantle, Australia, plans to have the system — the latest attempt at harvesting power from the sea — up and running by June. The pilot project will generate a lot of press, but veterans of the marine-energy field will watch warily to see how it fares. This industry has taken a slow road: none of the myriad devices designed thus far has proved its worth in the highly competitive energy market, and few have survived prolonged exposure to the harsh conditions in the sea. Despite an overall investment of around US$735 million over the past decade by a dozen leading companies, marine power from tide and waves has yet to take off. In fact, it remains the most expensive form of power on Earth.

But the outlook has brightened for those hoping to tap this source of energy. In the past few years, several major industrial leaders have acquired start-up companies that harvest energy from tides, the easiest type of marine power to capture. In March, three projects were approved for Canada's Bay of Fundy, home to some of the largest tides on the globe. The wave-energy industry, which is targeting a much larger but more elusive resource, has had a few setbacks, including a decision to scale back plans for an array off the coast of Oregon last month. But there is little doubt that both types of marine energy will eventually grow. Last year, the London-based consultancy Bloomberg New Energy Finance projected that up to 22 tidal projects and 17 wave projects generating more than one megawatt of electricity — enough to power around 250 homes — could be installed by 2020.

In theory, oceans could power the entire globe without adding any

© 2014 Macmillan Publishers Limited. All rights reserved


SIEMENS AG

Marine Current Turbines' tidal-energy converter in Strangford Lough, UK, generates power with underwater blades that can be raised for maintenance.

pollution to the atmosphere. And they could provide a more dependable source of electricity than the wind or sun. They are also geographically convenient: roughly 44% of the global population lives within 150 kilometres of the coastline.

Some nations have been using large, dam-like structures to block off inlets and draw power from tidal flow for decades, but the latest approaches are designed to be less intrusive. Although the potential environmental impacts are still under investigation, many researchers say that the sea could turn out to be an even more benign source of power than wind.

Indeed, energy experts envisage a day when the sea will deliver a significant amount of reliable carbon-free power to islands and burgeoning coastal cities around the world. "It has proved harder than people had expected at the start, but it has also proved possible," says Neil Kermode, managing director of the European Marine Energy Centre, the leading test facility for wave- and tidal-energy devices in the Orkney Islands, UK. "We have shown that you can make electricity from moving sea water, and that's a huge step forwards."

Shifting tides

Twice a day, some 350 million cubic metres of tidal water flows through a narrow strait into Strangford Lough — a small inlet southeast of Belfast, UK — and then back out to sea. With each trip, it passes a pair of 16-metre-long propellers attached to a central tower that is anchored to the floor of the channel. The force of the water, equivalent to a wind blowing at 555 kilometres per hour, spins the propellers at up to 15 revolutions per minute, generating 1.2 megawatts of electricity.

In addition to the traditional propeller, tidal-power companies have experimented with contraptions such as corkscrews, hydrofoils and underwater kites. The device used in Strangford Lough, however, is leading the way. Built by Marine Current Turbines in Bristol, UK, the design has generated more than 90% of the industry's power to date, according to the company.

The achievements attracted the interest of engineering giant Siemens in Munich, Germany, which took control of the company in 2012. Marine Current Turbines is now preparing to deploy its first array of five 2-megawatt machines — each costing roughly £9 million (US$15 million) — off the coast of Wales by 2016. As well as boosting the size of the machines, the company added a third blade to reduce vibrations and make the machines more durable, says chairman Kai Kölmel. But he cautions that progress is likely to be incremental. "I think some of the venture capitalists are disillusioned, but this is not a quick-buck industry," he says. "The wind industry didn't start up rapidly either."

Even with major companies such as Siemens entering the game, the biggest challenge remains attracting the money needed to do the engineering and build prototypes, says Christopher Sauer, chief executive of Ocean Renewable Power Company, based in Portland, Maine. Sauer's company has developed and deployed, albeit briefly, a unique device off the coast of Maine that looks a bit like the spinning blades of a combine harvester. The company is now working on a second-generation device that should be ready for deployment as early as 2015. "We're doing the best that we can with the money we have," he says.

Waves of energy

The power of waves is vast, but developing machines that can reliably extract that energy and withstand the often-punishing environment represents a wholly different kind of challenge. Companies have explored designs ranging from swinging flaps to gyroscopic devices that convert the rocking of a ship into a circular motion to drive an onboard generator. Each has its advantages, but the idea behind Carnegie Wave Energy's bobbing buoys in Australia was to escape the rough battering of waves at the surface. The submersion has the added advantage of keeping the devices out of sight and avoiding the debates over aesthetics that have arisen with wind farms.

As the waves carry the buoys up and down, the sea-floor pumps circulate fluid through a closed loop that extends roughly 3 kilometres to an onshore generation facility (see 'Water works'). Operating like a bagpipe, the system accumulates pressure and then releases it gradually to generate a constant flow of electricity. Each of the three devices can generate up to 240 kilowatts of energy.

"This is obviously not a commercial project, but there's not a commercial wave-energy facility in the world at the moment," says Carnegie's chief operating officer Greg Allen. Nonetheless, it represents progress, he says: each device generates three times more power than the version that was tested in the same waters in 2011. Allen says that the first commercial projects could come as early as 2018. To gain a toehold in the market, the company aims to compete with diesel-fired power generators on islands.

The sea could turn out to be an even more benign source of power than wind.

Pelamis Wave Power, based in Edinburgh, UK, has taken a different approach. It uses a set of five connected buoys that float on top of the ocean and wriggle in the waves like a snake. The segments move independently, and hydraulic pumps at each of the joints use the motion to drive fluids to an on-board generator. The company is currently testing a pair of 750-kilowatt machines at the Orkney Islands testing site. One of these is operated through a partnership with Scottish Power Renewables, a utility based in Glasgow. New components have helped to reduce internal wear and tear in the hydraulic system, and the company is now working on algorithms that will allow the device to adjust each of its 16 hydraulic pumps individually and maximize energy production as it rides the waves.

The company hit some rough waters last year when the German energy firm E.ON pulled out of a partnership after a three-year project to test one of the devices in the Orkney Islands. Founder and chief executive Richard Yemm remains optimistic, and says that the device is still in the water and producing power. But he acknowledges that the company needs to strengthen its bonds with the industrial energy sector. "We need the big industrial skills that the big companies have," he says.

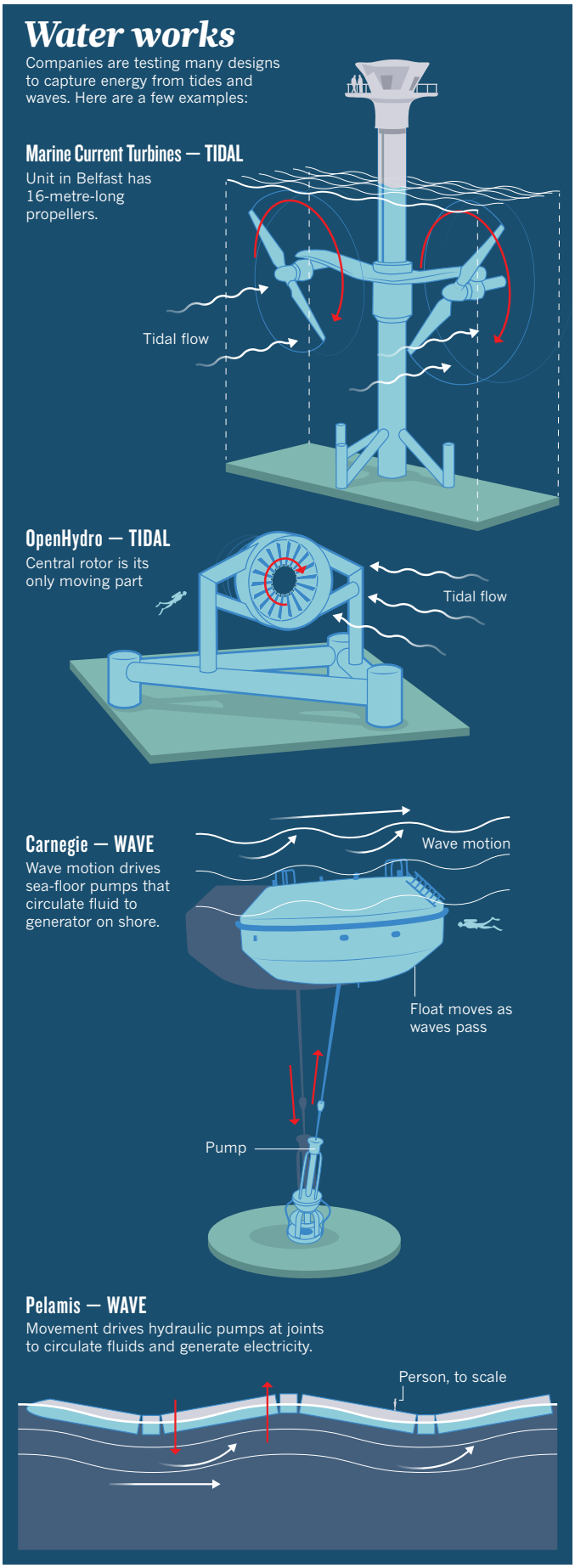
One reason that wave energy has attracted less commercial interest is that no one has yet managed to produce machines that can withstand the rough and tumble of the seas and produce electricity around the clock for sustained periods, says Angus McCrone, chief editor at Bloomberg New Energy Finance. "These companies have gone through a lot of money," he says, "and now the industry finds that it is going to have to spend significantly more to get the devices to commercialization."

Green machines

The marine-energy industry has also encountered wary regulators who are well aware of controversies surrounding bird deaths caused by wind turbines. Before Marine Current Turbines was allowed to start testing, it had to place seal spotters atop its turbines, who would hit an emergency cut-off button if any seals approached (which did not happen). And fears that seabed-mounted turbines designed by OpenHydro in Dublin, Ireland, could turn orcas into whale sushi nearly killed a proposal to test the devices in the Puget Sound near Seattle, Washington.

Gayle Zydlewski, a fish biologist at the University of Maine in Orono, says that she was able to capture only limited data on fish moving around and through Ocean Renewable Power's tidal unit in Maine. Fish mostly seemed to avoid the turbine, she says, but she wonders: "What happens when you have another one next to it?" Her team is still collecting baseline data, and her goal is to improve her models and understand

© 2014 Macmillan Publishers Limited. All rights reserved



NIK SPENCER/*NATURE*

how much fieldwork will be needed to determine potential impacts.

Others are busy in the lab. Biologists with the US Department of Energy have shot fish through turbines and exposed them to electromagnetic fields similar to those that surround the cables that transport the power back to shore. Neither seemed to cause any permanent damage, according to the department. In the case of the orcas in Puget Sound, energy-department researchers at the Pacific Northwest National Laboratory in Richland and Sandia National Laboratories in Albuquerque, New Mexico, studied a worst-case scenario: what if a curious orca stuck its head inside one of the turbines? The teams scoured the literature, tested various rubber materials as surrogates for whale skin and built a model to understand the potential impact of a blade strike (see go.nature.com/aaptn2). When a dead whale washed up onto the shore near Seattle last year, the scientists performed a computed tomography scan on the skull to look at skin and blubber thickness and used the information to improve their models. They also took some whale skin and tested its strength in the lab.

The upshot of the study, released in January, was that an orca that got struck on the head by a turbine blade would probably swim away with little more than a bruise, says biological oceanographer Andrea Copping from Pacific Northwest National Laboratory, who led the effort. "When whales hit a ship, it's the jawbone break that causes the death, but there wasn't enough force for that to happen," she says. The proposal to test the turbines in Puget Sound was approved by the Federal Energy Regulatory Commission on 20 March.

Copping is also heading an international collaboration that is pulling together all the environmental research on tidal- and wave-energy development. The goal is to determine the most likely impacts and concentrate the research on those.

Its first report, issued in January 2013, focused on three areas: animal interactions, turbine noise and the effect of taking energy out of the marine system and slowing down the flow of water. Thus far, the team reports, there is no evidence that there would be significant effects on marine animals or water flow, although the influence of large arrays is difficult to project.

The issue of sound could be more difficult to resolve. Researchers have taken detailed measurements of individual devices, and blasted fish at volume levels that the animals would experience if they were effectively strapped to the turbines for 24 hours. Apart from some tissue damage, equivalent perhaps to what a teenager would experience at a rock concert, Copping says that the fish seem to be okay. But the broader effects of an array of devices in an environment that is already saturated with sound from the motion of water as well as the loud drone of ships are hard to forecast. A little noise might help the animals to avoid the machines, but too much could trouble whales and other animals that use sound to communicate."Many or all of these projects are going to need a good deal of monitoring," Copping says. "The oceans are everybody's backyard."

Developers, researchers and environmentalists all agree on one thing: to understand both the economics and the environmental impacts, the industry needs to put more machines in the water. Bloomberg's McCrone says the projections for wave power are likely to drop in their next assessment, owing to declining industrial interest and the cancellation of some projects. But he also believes that both sectors will eventually mature.

One of the current hot spots is Canada's Bay of Fundy, which will soon host three projects, including a 4-megawatt installation featuring two devices from OpenHydro that will produce enough power for 1,000 homes by 2015. If all goes well, the company hopes to expand the array to eventually generate 300 megawatts. Although that would equal just one small coal-fired electricity plant, it would be a significant step forward for marine power.

"It will take off eventually," McCrone says. "There's a lot of energy in the sea." ■

Jeff Tollefson *writes about energy and the environment for* Nature.

© 2014 Macmillan Publishers Limited. All rights reserved



THE PLASTICS PUZZLE

When toxicologists warned that the plastics ingredient BPA might be harmful, consumers clamoured for something new. But problems persist.

BY JOSIE GLAUSIUSZ

© 2014 Macmillan Publishers Limited. All rights reserved

A stroll down the aisles of a US supermarket reveals a modest victory for consumer activism. In the baby-products section, plastic baby bottles, spill-proof cups and miniature cutlery are proudly marked 'BPA-free' — a sign that they no longer contain the compound bisphenol A, found in many plastics. A range of blenders and water bottles in the kitchenware aisle are also untainted by the chemical, as are a few cans of beans tucked away in the organic foods section. And customers filling their baskets with these BPA-free treasures may even receive a BPA-free receipt at the cash register.

The partial withdrawal of BPA is the culmination of two decades of research and hundreds of studies linking the compound — which mimics sex hormones called oestrogens — to adverse health effects in rodents and humans. The decision by regulators in the United States and European Union to ban BPA from baby bottles, combined with industry marketing campaigns, has convinced many consumers that the plastics and other containers currently used to store food are safe.

It is a false sense of security. BPA is still a constituent of many food containers, especially cans. And when companies did abandon BPA, they often adopted compounds — such as the increasingly common bisphenol S (BPS) — that share much of the same chemistry and raise many of the same concerns as BPA. "People use this chemical to replace BPA without sufficient toxicological information," says Kyungho Choi, an environmental toxicologist at Seoul National University. "That is a problem."

FALL FROM GRACE

BPA has formed the chemical backbone of most hard, clear polycarbonate plastic since the 1950s. Over time, studies have linked the chemical — which can leach out of plastics and into food — to a host of adverse health effects, including reductions in fertility and birth weight, male genital abnormalities, altered behavioural development, diabetes, heart disease and obesity[1] (see *Nature* **464,** 1122–1124; 2010).

Establishing a clear connection between a compound such as BPA and human health is complex, notes Geoffrey Greene, who studies oestrogens and their receptors at the University of Chicago in Illinois. "Most studies address only the question of whether such adverse effects can occur in various cell- or animal-based models," he says, "without addressing whether the amounts that we are exposed to are sufficient to have an effect on human health." Because many of the potential health effects of BPA are difficult to assess, the US National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina, has launched a US$30-million research programme designed to answer outstanding questions.

A few years ago, mounting evidence and concerned consumers convinced governments to take action. In 2011, the European Union banned BPA from baby bottles; the United States followed suit a year later. But BPA-based linings are still slathered on the insides of most food and beverage cans, and used to coat water-supply pipes in many countries. The compound is also found in dental sealants and in incubators for premature infants.

In the quest to replace BPA, finding an alternative for food and drink cans has proved particularly vexing. Creating a cheap lining for tins suitable for a range of foods — from beans to tomatoes to haggis curry — is no simple matter. Not only must the packaging prevent bacteria and fungi from attacking the food, the can's lining must also stop the food from attacking and corroding the can. Moreover, when metal comes into contact with food, it can ruin the flavour. "If your food tastes funny, but you tell people it's safer, are they going to believe that?" says Daniel Schmidt, a plastics engineer at the University of Massachusetts in Lowell.

Manufacturers also prefer linings that block sulphur compounds — found in proteins, preservatives and pesticides — from reacting with the metal and forming unattractive iron or tin sulphide stains. No BPA-free lining has yet emerged that can accomplish all of this. "You have to have a special can and a special coating for every class of food," Schmidt says. "It gets extremely cumbersome and quite expensive."

BPA-based epoxy linings are widely used because they are strong, flexible and cheap. They tolerate the high temperatures needed to sterilize foods during canning, and do not interact with a huge array of foods and beverages, according to the North American Metal Packaging Alliance in Washington DC. The alliance estimates that 95% of all aluminium and steel can coatings are epoxy-type resins: more than 99.9% of these contain BPA.

And although there are alternatives, these are not without drawbacks. In 1999, Eden Organic of Clinton, Michigan, began coating bean tins with plant-based oleoresins. The switch has increased the cost of the tins by more than 20%. Oleoresin linings can also alter the taste of food, and are vulnerable to attack from high-acid foods — such as tomatoes — which have a particular propensity to leach BPA, according to biologist Frederick vom Saal of the University of Missouri-Columbia.

Some Japanese manufacturers now use reduced-BPA lacquers to coat cans. Other coatings include acrylics, which are too brittle for use in many tins, and vinyls and phenolics, both of which may have oestrogenic effects.

New options are beginning to surface. Schmidt is developing an epoxy based on a molecule in Tritan — a BPA-free polymer manufactured by the Eastman Chemical Company in Kingsport, Tennessee, that is used in baby bottles. He hopes that his epoxy will be as versatile and cheap as BPA. "The profit margins are so thin in the can-coating industry," he says. "Unless somebody hands them that solution, it's going to be tough for them to accept anything."

In contrast to the can-liner conundrum, replacing BPA in baby bottles and cash-register receipts proved relatively straightforward. When BPA fell from grace, many manufacturers turned to the compound's structural kin: BPS. A BPA molecule consists of two phenol groups connected by a branched three-carbon group. In a BPS molecule, the two phenol groups are instead connected by a sulphone group (SO_2).

BPS was first made in 1869 as a dye. But because it was introduced into consumer goods only recently — into cash-register receipts in 2006, for example — few researchers have studied its toxicity. "The main question, to which we have no answer, is: 'is BPS as toxic as BPA?'" says René Habert, an endocrinologist at Paris Diderot University.

A KEY TO ACTION

The similarity of BPS's structure to that of BPA is enough to raise suspicions that it may mimic oestrogens, says Cheryl Watson, a biochemist at the University of Texas Medical Branch in Galveston. Natural oestrogens are small molecules containing several phenolic rings; these bear chemical adornments that bind to a pocket found in oestrogen receptors in the body. BPA and BPS are about the same size and have similar phenolic rings with similar attachments, so they may slot like keys into oestrogen receptors, Watson says.

Watson and a colleague, Rene Viñas, now at the US Food and Drug Administration, measured the responses of cultured rat pituitary cells to BPS. These cells are particularly sensitive to oestrogens and oestrogen mimics, allowing the team to study concentrations of BPS down to 10^{-15} moles per litre. The team found that even at these very low levels, BPS triggered the enzyme cascade normally activated by an oestrogen called oestradiol[2], an effect also seen with BPA. When combined with levels of oestradiol found in adult women, BPS seemed to over-stimulate the pathway, shutting it down and causing cell suicide. The results, says Watson, were typical of those expected of an oestrogen mimic: inappropriate activation of oestrogen responses, disruption of normal oestrogen-response pathways, and eventual cell death.

Others have seen similar effects. Susanne Bremer and her colleagues at the Institute for Health and Consumer Protection, a European Commission-funded research centre in Ispra, Italy, tested BPS and BPA on an oestrogen-sensitive human cell line. They found that both chemicals behaved like oestrogens, but were 100,000-fold less active than oestradiol[3]. Choi and his colleagues discovered that zebrafish exposed to 0.5 micrograms of BPS per litre of water — about one-sixth of the maximum concentration detected in the environment — had fewer eggs, more malformed offspring and higher oestrogen to testosterone ratios than untreated zebrafish[4]. "High concentrations of BPS have the same effect as high concentrations of BPA," says Habert, who has conducted

PAVEL L. PHOTO AND VIDEO/SHUTTERSTOCK.COM

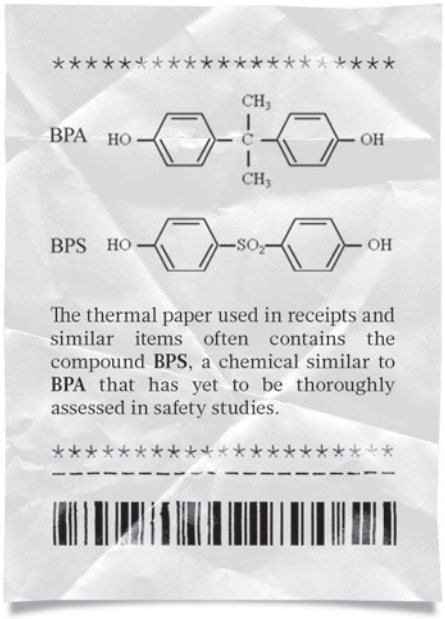
© 2014 Macmillan Publishers Limited. All rights reserved

preliminary experiments on the effect of BPS on mouse and human fetal testis cells. "At low concentrations, the effect is unknown."

What concentration best approximates human exposure to BPS is not clear. A team led by Catherine Simoneau of the Institute for Health and Consumer Protection analysed a total of 30 BPS-containing baby bottles from 12 countries. After five minutes in boiling water and two hours at 70 °C, none of the bottles released detectable quantities of BPS[5]. "These materials are far more resistant to hydrolytic breakdown than polycarbonate — that was one of the big selling points," says Schmidt. "As such, I would consider them to be safer than polycarbonate in a food-contact setting."

But people are exposed to BPS in many different ways. Kurunthachalam Kannan, an analytical chemist at the New York State Department of Health in Albany, and his team have found BPS on cash-register receipts, and aeroplane luggage tags and boarding passes, all of which are made from thermal paper containing BPS as a colour developer. The scientists also found BPS in products made from recycled paper, including pizza boxes and food buckets.

Kannan's team estimates that the average daily exposure to BPS through the skin is well below the threshold values for toxic effects. Nevertheless, given the potential for higher levels of exposure from other sources such as food, Kannan urges further studies of the compound. And Watson argues that even small amounts of oestrogen mimics can cause trouble. "The problem is that they are active in such small quantities," she says. "If you leach even a little, you still leach enough for responses to happen."



The thermal paper used in receipts and similar items often contains the compound **BPS**, a chemical similar to **BPA** that has yet to be thoroughly assessed in safety studies.

BRANCHING OUT

Some manufacturers have left the bisphenol family in search of a replacement. In 2007, the Eastman Chemical Company launched Tritan — a new heat-resistant clear plastic — for infant-care products such as baby bottles. This BPA-free plastic has since replaced the old BPA-containing polycarbonate in many water bottles, food containers and children's cups. Eastman says that the results of testing, analysed by Thomas Osimitz of Science Strategies, a consulting firm in Charlottesville, Virginia, and his colleagues, verified that Tritan's monomers do not bind to oestrogen or androgen receptors[6].

In 2011, George Bittner, a neurobiologist at the University of Texas at Austin and the chief executive of Austin-based chemical-testing company CertiChem, reported that 92% of 102 commercially available plastic products leached chemicals with oestrogenic activity[7]. This included plastics advertised as BPA-free. The reason, Bittner says, is that additives in plastics — such as stabilizers and lubricants — can also bind to oestrogen receptors, as can some of the plastic monomers themselves.

Tritan resins produced by Eastman were among the polymers that showed oestrogenic activity in Bittner's assays. When PlastiPure, a sister company of CertiChem, produced a brochure publicizing these results, Eastman sued. The company's lawyers maintained that the *in vitro* test used by CertiChem — which involved oestrogen-sensitive breast cancer cells grown in culture — is not a definitive assay for oestrogenic activity. In a letter to the editor of the journal *Food and Chemical Toxicology*, Bittner countered that his assays are up to 200 times more sensitive than the tests Osimitz analysed to demonstrate Tritan's safety[8].

Bittner had support: Wade Welshons, who studies endocrine disruptors at the University of Missouri-Columbia, independently tested five Tritan bottles using the same assay as Bittner. In a deposition entered during the trial, Welshons reported that he found detectable oestrogenic activity in each test. But the jury ruled in Eastman's favour, and the judge barred Bittner, PlastiPure and CertiChem from making claims about Tritan's oestrogenic activity.

Eastman stands by the results reported by Osimitz. And, unlike Bittner, Welshons suspects that the oestrogenic activity he found is attributable not to the Tritan polymer itself, but to other compounds added during plastics manufacturing. He is not the only one concerned about the complex mixtures of chemicals used to make plastics. In 2012, the world produced some 280 million tonnes of plastic. According to a model based on the United Nations' Globally Harmonized System of Classification and Labelling of Chemicals, more than 50% of these plastics contain ingredients that can be hazardous (see *Nature* **494,** 169–171; 2013). Some are carcinogenic; others are oestrogenic.

It is not yet clear how many of these chemicals are dangerous at the concentrations found in the plastics. But mixed together, the chemicals could have synergistic effects. Watson and Viñas recently studied the effect of the oestrogen mimics BPA, BPS and nonylphenol (a detergent precursor) on cultured rat pituitary cells. They found that a combination of two or three of the compounds caused greater disruption to the oestrogen-signalling system — and did so at lower concentrations — than did a single compound[9]. "We don't experience any of these chemicals alone," Watson says. "A lot of other chemicals mimic oestrogens."

Ideally, says Watson, the next generation of chemicals would be tested for effects on oestrogen signalling before widespread deployment in food containers. To that end, she and a group of biologists and chemists have put together a plan called TiPED, or Tiered Protocol for Endocrine Disruption. Under this testing system, newly synthesized chemicals would be evaluated for endocrine-disrupting potential at five different stages, from initial computational analysis of structure to whole-animal experiments.

The goal is to form a consortium of independent laboratories that would test chemicals on request by plastics companies. Convincing these companies to participate will be a challenge, Watson acknowledges. But there is an incentive, she argues, because companies can suffer from bad press, lost business and lawsuits when a chemical they produce or use is linked to human health concerns. "When it gets proved that an individual chemical is a problem," says Watson, "they're going to have to, as an industry, completely retool."

The TiPED proposal is designed to ensure that endocrine-disrupting chemicals no longer reach the market. For Watson and many other researchers, the current situation raises concern because there are so many untested compounds found in countless plastic products. Those chemicals, she says, "are really all around us". ■

Josie Glausiusz *is a freelance journalist in Ra'anana, Israel.*

1. Rochester, J. R. *Reprod. Toxicol.* **42,** 132–155 (2013).
2. Viñas, R. & Watson, C. S. *Environ. Health Perspect.* **121,** 352–358 (2013).
3. Grignard, E. Lapenna, S. & Bremer, S. *Toxicol. in Vitro* **26,** 727–731 (2012).
4. Ji, K., Hong, S., Kho, Y. & Choi, K. *Environ. Sci. Technol.* **47,** 8793–8800 (2013).
5. Simoneau, C., Valzacchi, S., Morkunas, V. & Van den Eede, L. *Food Addit. Contam.* **28,** 1763–1768 (2011).
6. Osimitz, T. G. *et al. Food Chem. Toxicol.* **50,** 2196–2205 (2012).
7. Yang, C. Z., Yaniger. S. I., Jordan, V. C., Klein, D. J. & Bittner, G. D. *Environ. Health Perspect.* **119,** 989–996 (2011).
8. Bittner, G. & Yaniger, S. *Food Chem. Toxicol.* **50,** 4236–4237 (2012).
9. Viñas, R. & Watson, C. S. *Environ. Health* **12,** 26 (2013).

© 2014 Macmillan Publishers Limited. All rights reserved



Credit where credit is due

Liz Allen, **Amy Brand**, **Jo Scott**, **Micah Altman** and **Marjorie Hlava** are trialling digital taxonomies to help researchers to identify their contributions to collaborative projects.

Research today is rarely a one-person job. Original research papers with a single author are — particularly in the life sciences — a vanishing breed. Partly, the inflation of author numbers on papers has been driven by national research-assessment exercises. Partly, it is the emergence of big and collaborative science, assisted by technology, that is changing the research landscape.

What we cannot tell easily by reading a paper is who did what. That is difficult to decipher by consulting the author lists, acknowledgements or contributions sections of most journals; and the unstructured information is difficult to text-mine[1,2].

Developments in digital technology present opportunities to do something about this. With the right 'taxonomy', manuscript-submission software could enable researchers to assign contributor roles relatively easily in structured formats during the process of developing and publishing a paper. An analogy is the FundRef initiative developed by funders, publishers and manuscript-submission vendors to build direct links between published research and associated funding sources during manuscript submission.

For researchers, the ability to better describe what they contributed would be a more useful currency than being 'author number 8 on a 15-author paper'. Scientists could draw attention to their specific contributions to published work to distinguish their skills from those of collaborators or competitors, for example during a grant-application process or when seeking an academic appointment. This could benefit junior researchers in particular, for whom the opportunities to be a 'key' author on a paper can prove somewhat elusive. Methodological innovators would also stand to benefit from clarified roles — their contributions are not reliably apparent in a conventional author list[3–6]. It could also facilitate collaboration and data sharing by allowing others to seek out the person who provided, for example, a particular piece of data or statistical analysis.

Through the endorsement of individuals' contributions, researchers can start to move beyond 'authorship' as the dominant measure of esteem. For funding agencies, better information about the contributions of grant applicants would aid the decision-making process. Greater precision could also enable automated analysis of the role and potential outputs of those being funded, especially if those contributions were linked to an open and persistent researcher profile or identifier. It would also help those looking for the most apt peer reviewers. For institutions, understanding a researcher's contribution is fundamental to the academic appointment and promotion process.

Such a system could benefit publishers too. Many journals do issue strict guidelines for what constitutes authorship, although there have been calls to overhaul these to reflect the reality of today's research[7,8]. Greater transparency should help to reduce the number of authorship disputes being managed by journal editors, and should cut the time that editors spend chasing listed authors for confirmation of their roles.

CLASSIFYING CONTRIBUTION

To probe how such a taxonomy might work, we conducted an experiment. Our findings, which are summarized here, set the stage for the development of a system or process that could change how contributions to research output are valued.

In 2012, a small group of journal editors joined forces with Harvard University in Cambridge, Massachusetts, and the Wellcome Trust in London to develop a simple contributor role taxonomy to test with researchers[9]. Some journals, such as those published by the Public Library of Science (PLOS), have been working with basic contribution classifications for a couple of years; the group decided to extend this.

An online survey, live between August and November 2013, was designed to test whether authors' contributions to recent journal articles could be classified using a 14-role taxonomy (see 'Who did what?'). The survey was sent to 1,200 corresponding authors of work published in PLOS journals, Nature Publishing Group journals, Elsevier journals, *Science* and *eLife*. Corresponding authors were asked to indicate the contribution of each author of their article according to the roles in the taxonomy, and to comment on its comprehensiveness; whether there were any significant role descriptors missing; how using the taxonomy compares with current author-contribution assignment; and specifically, how easy or difficult it was to use.

Around 230 authors gave feedback. More than 85% found the taxonomy easy to use and felt that it covered all the roles of contributors to their paper. Furthermore, 82% of respondents reported that using the more-structured taxonomy of contributor roles presented to them was at least 'the same' as (37%) or 'better' (45%) in terms of accuracy than how the author contributions to their recently published paper had actually been recorded.

There is certainly more work to do. The pilot yielded substantial feedback on several themes. These included: how to ensure agreement among authors on their specific contributions; how to prevent supervisors from inappropriately taking credit; whether to distinguish between 'lead', 'supporting' and 'equal' roles; and how to recognize that the significance and relevance of certain roles varies between articles and research areas. Others suggested that more types of contribution should be included in the taxonomy or that some contributions such as 'funding acquisition' and 'project management' might be captured elsewhere in the manuscript-submission process.

There are also methodological caveats associated with this pilot: the sample was relatively small and only corresponding authors were asked for their opinions. The taxonomy was developed and tested in the biomedical and life-sciences community — we have not tested its validity in other fields because we

© 2014 Macmillan Publishers Limited. All rights reserved

WHO DID WHAT?
Respondents were asked to select all roles that applied to each author, as described in the taxonomy below, and to state which of these roles were lead or supporting.

Taxonomy category	Description of role
Study conception	Ideas; formulation of research question; statement of hypothesis.
Methodology	Development or design of methodology; creation of models.
Computation	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms.
Formal analysis	Application of statistical, mathematical or other formal techniques to analyse study data.
Investigation: performed the experiments	Conducting the research and investigation process, specifically performing the experiments.
Investigation: data/evidence collection	Conducting the research and investigation process, specifically data/evidence collection.
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation or other analysis tools.
Data curation	Management activities to annotate (produce metadata) and maintain research data for initial use and later re-use.
Writing/manuscript preparation: writing the initial draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft.
Writing/manuscript preparation: critical review, commentary or revision	Preparation, creation and/or presentation of the published work, specifically critical review, commentary or revision.
Writing/manuscript preparation: visualization/data presentation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.
Supervision	Responsibility for supervising research; project orchestration; principal investigator or other lead stakeholder.
Project administration	Coordination or management of research activities leading to this publication.
Funding acquisition	Acquisition of the financial support for the project leading to this publication.

expect that there are field-specific contributor roles. Nonetheless, this feedback provides a springboard to further explore how a system for allocating contributor roles might be implemented.

SETTING STANDARDS

So what now? Over the next six to eight months, we will be collaborating with bodies such as the National Information Standards Organization to evolve the taxonomy. Through this collaboration, we will consult a broader cross-section of the research community, including researchers from different scientific fields, to see how valid the taxonomy might be beyond biomedicine, and to ascertain the value that greater definition of contributor roles would bring to the research ecosystem. We are mindful that any approach must not add to researchers' burdens in submitting and publishing work, or fuel authorship dissatisfaction. For instance, one respondent described our trial taxonomy as: "more accurate and less 'generous'".

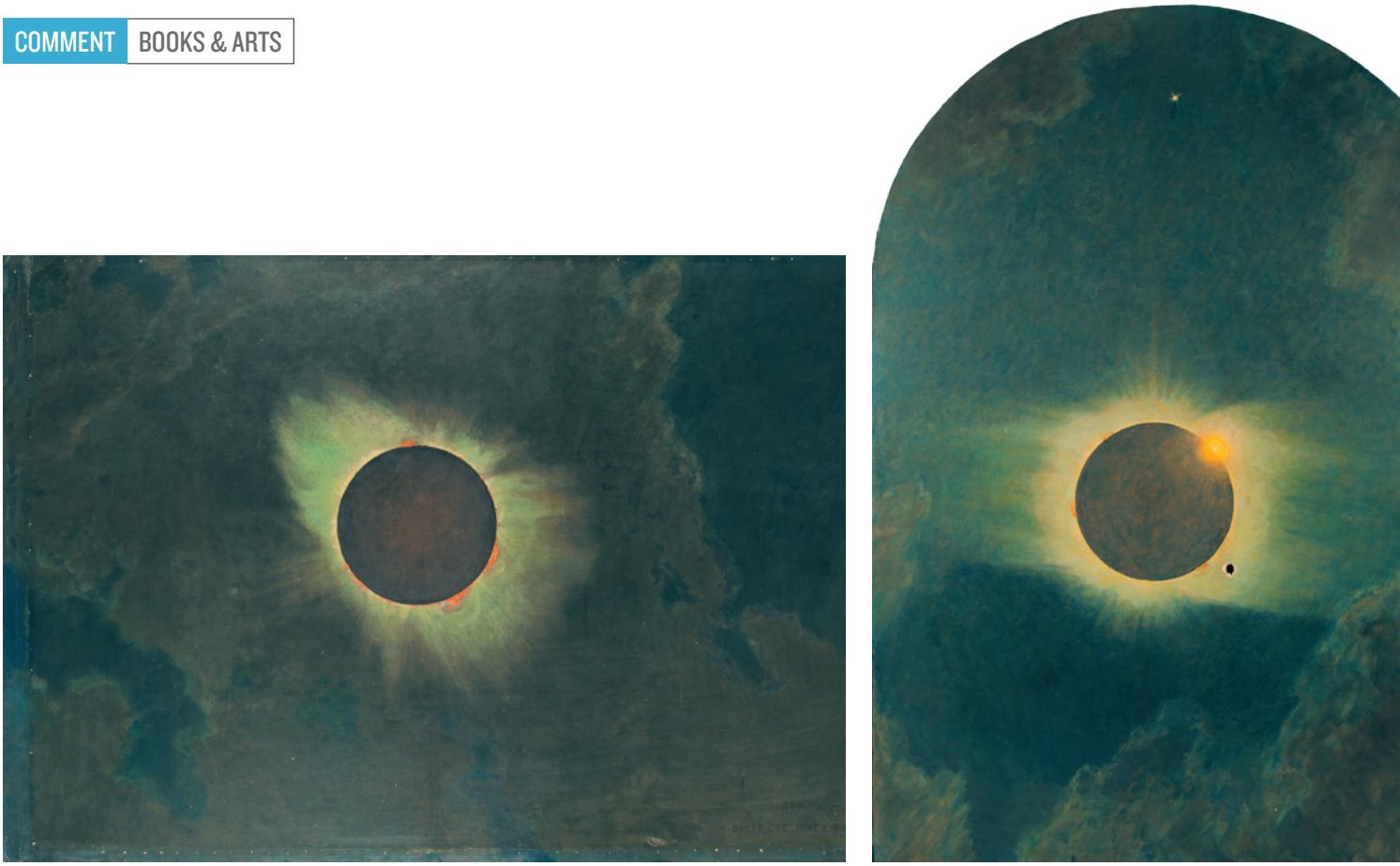
A second workshop on contributor roles is planned for the third quarter of 2014, after which we intend to implement a fuller trial across more research publication outlets and disciplines in 2015. Models of implementation to be tested could include, for example, integrating a digital taxonomy with manuscript-submission and research-management systems. The latter approach imagines assigning and agreeing on contributor roles before preparing the manuscript for publication, with the potential to profoundly affect the culture and process of doing research. ■



Liz Allen *and* **Jo Scott** *are at the Wellcome Trust in London, UK.* **Amy Brand** *is at Digital Science in Cambridge, Massachusetts, USA.* **Marjorie Hlava** *is at Access Innovations in Albuquerque, New Mexico, USA.* **Micah Altman** *is at the MIT Libraries, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.*
e-mail: l.allen@wellcome.ac.uk

1. Dance, A. *Nature* **489,** 591–593 (2012).
2. Frische, S. *Nature* **489,** 475 (2012).
3. Karani, R., Ognibene, F. P., Fallar, R. & Gliatto, P. *Acad. Med.* **88,** 364–368 (2013).
4. Rodrigo, C. & Bordons, M. *Scientometrics* **88,** 145–161 (2011).
5. Cleary, M., Jackson, D., Walter, G., Watson, R. & Hunt, G. E. *J. Clin. Nurs.* **21,** 809–811 (2012).
6. Einav, L. & Yariv, L. *J. Econ. Pers.* **20,** 175–187 (2006).
7. Matheson, A. *PLoS Med.* **8,** e1001072 (2011).
8. Malički, M., Jerončić, A., Marušić, M. & Marušić, A. *BMC Med. Res. Methodol.* **12,** 189 (2012).
9. Hames, I. *Report on the International Workshop on Contributorship and Scholarly Attribution* (Harvard University, Wellcome Trust, 2012).

© 2014 Macmillan Publishers Limited. All rights reserved



Howard Russell Butler's triptych of solar eclipses, painted in 1918, 1923 and 1925: the first works by an artist to depict the solar corona accurately.

ASTRONOMY

Art of the eclipse

As the next solar eclipse approaches, **Jay M. Pasachoff** and **Roberta J. M. Olson** ponder how artists from the early Renaissance onwards have interpreted the phenomenon.

State-of-the-art photography can capture a solar eclipse in all its evanescent glory — as will be seen on 29 April, when the first solar eclipse of 2014 will be visible from Australia and Antarctica. But long before such technology existed, artists from the fourteenth to the early twentieth century portrayed this fleeting phenomenon in paint with increasing accuracy, on the basis of direct observations, scientific documents and contemporary theory. These are rare artworks: the phenomenon appears in one location on average roughly every 300 years, and its dramatic phases can be seen for just minutes from a narrow band of the planet hundreds of kilometres across, although thousands of kilometres long.

In the Middle Ages and even the early Renaissance, artists couched portrayals of eclipses in religious contexts redolent with symbolism. The Crucifixion was believed to have taken place during a total solar eclipse, and a few early renderings show a stylized, occulted Sun on one side of the cross. In the fourteenth century the Florentine artist Taddeo Gaddi, a student of Giotto, took a revolutionary step. On the inside shutter of a Crucifixion triptych, he suggested the darkened sky and strange light of a solar eclipse by painting a dark-blue wedge in one corner, rimmed faintly with now-tarnished silver. And in his frescoed *Annunciation to the Shepherds* in the Basilica of Santa Croce in Florence, Italy, Gaddi represented divine radiance using light effects that he had seen during the eclipse of 16 July 1330 — an observation that left him partially blind. He effectively conveys the eerie, quasi-nocturnal illumination.

Some 200 years later, in the High Renaissance, an artist captured the solar corona — the pearly halo surrounding the Moon's silhouette during total eclipse. In a loggia of the Vatican's Apostolic Palace around 1518–19, Raphael and his workshop frescoed a depiction of a solar eclipse with a dark lunar disk ringed by the innermost corona in *Isaac and Rebecca Spied Upon by Abimelech*. Although lacking details such as the coronal 'streamers' jutting out from active regions on the Sun, and otherwise showing some artistic licence, the spiky rays around the inner corona are roughly representative. No corona would have been visible during the 8 June 1518 annular eclipse that passed through Rome, but the event may have inspired Raphael.

An even more accurate portrayal was achieved another two centuries later, when Cosmas Damian Asam painted his *Vision of St Benedict* in 1735 for a monastery in Weltenburg, Bavaria. In the eclipse that confronts the elderly saint, light bursts from the edge of the lunar disk after totality — the first relatively faithful representation of the 'diamond-ring effect', which occurs when the first ray of light breaks through a valley on the Moon's edge. Asam must have seen the eclipse of 13 May 1733, and might also have consulted descriptions from contemporary scientists such as Edmond

NATURE.COM
For more on eclipses and the arts:
go.nature.com/ktxscq

© 2014 Macmillan Publishers Limited. All rights reserved



AMNH LIB.

Halley, whose predictions of eclipses had circulated some years before.

A handful of subsequent eclipse 'portraits' stand out, such as Ippolito Caffi's oil painting *View of Venice with the Eclipse of 8 July 1842*, which depicts the moment just before or after totality. Caffi shows one-quarter of the sky brightly lit and three-quarters of it dark, which is highly inaccurate; yet he does show an eclipse as a process involving dynamic changes in light.

Complete accuracy came towards the end of the twentieth century, when photography and computer analysis finally captured the nuances and phases of eclipses. But in that century's first decades, one artist had already achieved precision, and beauty, in paint.

PHOTO SCALA, FLORENCE



The corona at an eclipse is visible in Raphael's *Isaac and Rebecca Spied Upon by Abimelech.*

Howard Russell Butler was a high-profile portrait painter who had degrees in physics and law, and had studied painting in Mexico under Frederic Edwin Church, a US landscape painter obsessed with portraying astronomical phenomena. That mingling of science and art made Butler's professional life unusual and rich. In his 1923 book *Painter and Space, or The Third Dimension in Graphic Art*, for instance, he brought his knowledge of physics to bear on the realistic rendering of modelling, light and perspective in art. But it was his astronomical paintings that showed this duality most dramatically.

Highly methodical, Butler kept shorthand notes on spatial and colour details for his paintings. This was to prove invaluable when, at age 62, he was invited to join the US Naval Observatory for its expedition to Baker, Oregon, to view the total solar eclipse of 8 June 1918. Butler later noted that as a portraitist, "I generally asked for 10 sittings of 2 hours each. But all the time they would allow me on this occasion was 112 1/10 sec. As it turned out I got a trifle more, for their calculations were short by 3/10 of a second." Along with an oil painting, Butler reproduced his studies of this eclipse the following year in 'Painting the Solar Corona', an article in *Natural History*, the magazine of New York's American Museum of Natural History (AMNH).

Butler refined his methods over time, painting the eclipse of 1923 in Lompoc, California, and of 1925 in Middletown, Connecticut. The three artworks are the first to show the actual detailed shapes of the corona over the sunspot cycle. Butler created a vast, stunning triptych from them, installed in 1935 at the AMNH's newly established Hayden Planetarium, where it engaged hordes of visitors until the original building was torn down in 1997. The triptych is now in storage. Butler also created a half-size replica for the Fels Planetarium in Philadelphia, Pennsylvania, still hanging; one for New York State's Buffalo Museum of Science; and one newly rehung for long-term display at Princeton University in New Jersey.

Butler went on to paint the 1932 solar eclipse, and filled large canvases with close-ups of solar prominences — beautiful and complex shapes of gas shining in the red light of hydrogen above the solar limb, or Sun's edge, which are still studied in astronomy. Six of Butler's paintings remain in storage at the AMNH, along with some of his other works, such as fanciful views of Mars from vantage points on its moons.

Today, eclipse photography is highly advanced. Because the corona diminishes in brightness by a factor of more than 200 from the solar limb to a distance of one solar radius, filters radially graded in density are sometimes used. This allows film or electronic detectors to capture the corona's appearance in a single exposure. Computer scientists or astronomers can also combine dozens of photographs taken with different exposures to reveal the coronal structure, and to artificially enhance contrast to make its features stand out.

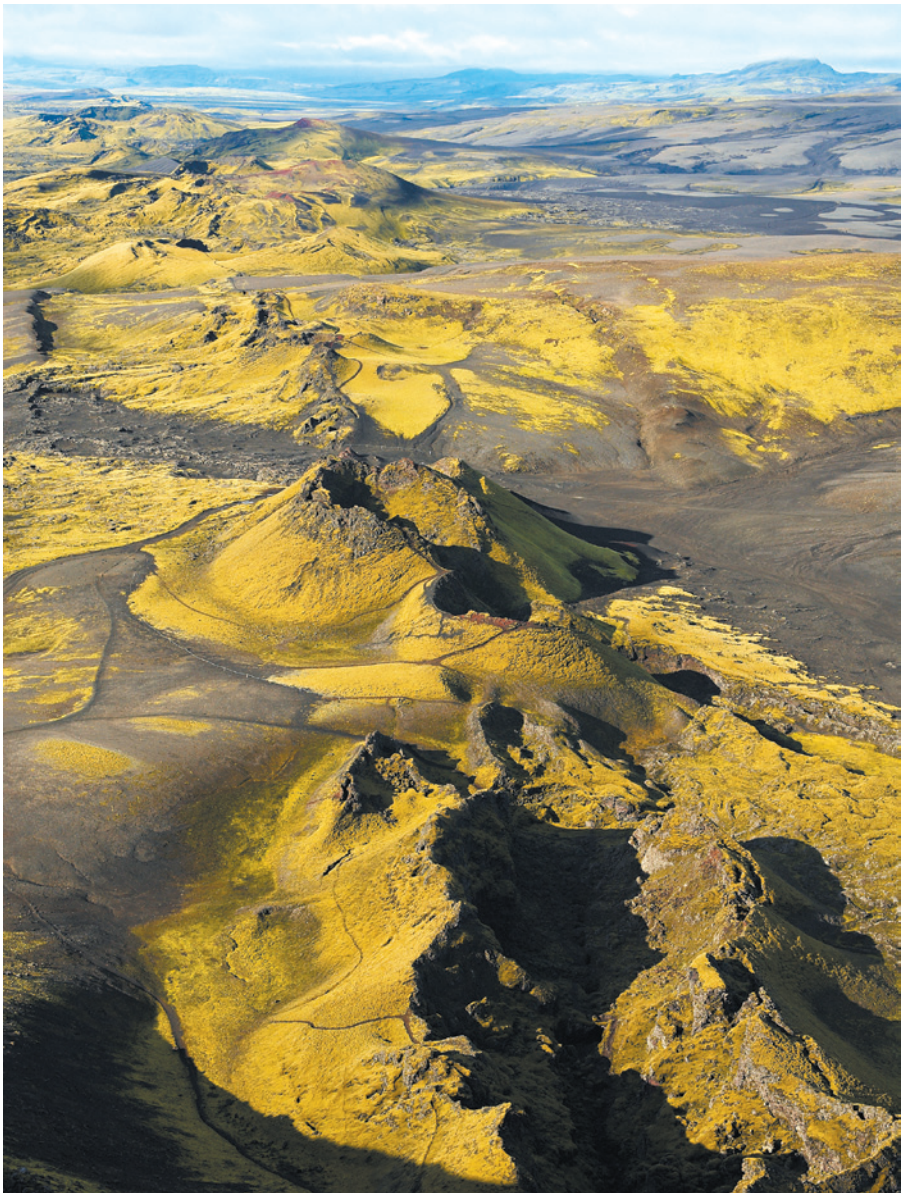
"Taddeo Gaddi represented divine radiance using light effects that he had seen during the eclipse of 16 July 1330."

But Butler's paintings, like Gaddi's, demonstrate that the acute visual perception cultivated by the handful of representational artists keen to probe this astronomical phenomenon enabled them to make startling observations, and even discoveries. We hope that the run-up to the 2017 total solar eclipse, which will cross the continental United States, will inspire a campaign to restore and display Butler's fascinating astronomical oeuvre. ■

Jay M. Pasachoff *is Field Memorial Professor of Astronomy at Williams College in Williamstown, Massachusetts, and chair of the American Astronomical Society's Historical Astronomy Division.*
Roberta J. M. Olson *is curator of drawings at the New-York Historical Society in New York City and Professor Emerita of Art History of Wheaton College in Norton, Massachusetts.*
e-mail: eclipse@williams.edu; rolson@nyhistory.org

© 2014 Macmillan Publishers Limited. All rights reserved

BRUNO MORANDI/HEMIS/CORBIS



Cinder cones in Skaftafell National Park, Iceland, mark the site of the 1783 Laki eruption.

GEOLOGY

Parsing eruptions

Ted Nield weighs up histories of two momentous volcanic events in Iceland and Indonesia.

Two chronicles of two volcanic eruptions — 32 years and thousands of kilometres apart — are emerging within months of each other. In *Island on Fire*, Alexandra Witze and Jeff Kanipe examine the late-eighteenth-century Lakagígar eruption (usually known as Laki) in Iceland. Gillen D'Arcy Wood gives us *Tambora*, a history of an 1815 colossal super-eruption in Indonesia that had a worldwide impact. Both left indelible marks in the historical and cultural, as well as geological, records.

Laki blew in 1783, when a section of southwestern Iceland split open and huge volumes of gassy lava and ash spewed out. The remains of the wall of fire can still be seen, striding across the landscape as a line of moss-covered cinder cones. The eruption lasted for eight months, and its global death toll is estimated in the millions. Crops failed across northern Europe, possibly hastening the onset of the French Revolution. Laki's distal effects were documented by Benjamin Franklin, then US ambassador in Paris, and British naturalist-priest Gilbert White, although neither knew for sure where the dry, sulphurous fogs came from. Similar eruptions are bound to recur, and in *Island on Fire* Witze and Kanipe pose two questions. What happened then? And how would we cope with such an event today, given the air-traffic chaos caused by the 2010 eruption of that relative pipsqueak Eyjafjallajökull?

Island on Fire: The Extraordinary Story of Laki: The Volcano That Turned Eighteenth-Century Europe Dark
ALEXANDRA WITZE AND JEFF KANIPE
Profile Books: 2014.

Tambora: The Eruption That Changed the World
GILLEN D'ARCY WOOD
Princeton University Press: 2014.

In describing Laki, Witze and Kanipe apply modern volcanological insight to a classic eyewitness account by Reverend Jón Steingrímsson (available in English as *Fires of the Earth*; Nordic Volcanological Institute, 1998). A scamper through plate tectonics and mantle-plume theory puts Iceland in context, before a survey of the world's 'supervolcanoes'. Subsequent chapters describe Laki's local and global effects, and revisit the site as it is now. A general tutorial on volcanic hazards precedes a final section on how modern society might fare, and when we can next expect to have to rise to that challenge. I need hardly shout 'spoiler alert' before I give away the unsurprising answers of 'not well', and 'possibly soon'.

The descriptions of modern interpretations of the eruption are lively and confident, and the book provides an excellent background that also broadens out to consider global volcanic hazards generally. But its didactic structure and approach can disappoint, coming as it does from two science writers (one of whom, Witze, works for *Nature*). The text resembles admirable lecture notes — touching on everything, with references to where one might find the more interesting details.

Despite this academicism, which many will welcome, the story is sometimes oversold. Europe was shrouded in dry fogs and suffered widespread acid rain, but it was never "turned dark" by Laki. Some geology is oversimplified and thereby rendered less interesting than it should be. Earth's magnetic field is not created by molten nickel and iron merely sloshing about in Earth's core, for example. The physics of the liquid dynamo are more controversial than Witze and Kanipe imply. Also given short shrift is Steingrímsson. The story of Laki hardly lacks harrowing tales of death by starvation, asphyxia or fluorine poisoning, much of it described by the cleric. Steingrímsson has long been revered — in life

© 2014 Macmillan Publishers Limited. All rights reserved

as a miracle-worker for seeming to stop the lava through the power of prayer when he celebrated a 'fire mass' that luckily coincided with the flow's final halt, and today as the meticulous chronicler of the eruption. His amazing career is here reduced to a litany of unfortunate events.

Tambora's stupendous 1815 eruption, by contrast, can hardly be oversold. Wood broadens our understanding beyond the 'year without a summer' cliché and the oft-parroted story of Mary Shelley and Lord Byron gloomily penning *Frankenstein* and *Darkness* in the eruption's aftermath while staying at the Villa Diodati near Lake Geneva, history's most famous holiday let. Tambora's eruption devastated the island of Sumbawa in Indonesia, and led to a decade of adverse weather that spread starvation and disease across North America, Europe, India and China — notably Yunnan Province, where successive crop failures turned a rice bowl into an opium-growing state.

Wood's command of the scientific literature is impressive, and more than matched by his knowledge of world history during this horrific episode of catastrophic global climate change. With the mass of information he has assimilated, he skilfully weaves a tale full of human and cultural interest, drawing in a "forgotten" 1816–18 Irish potato famine and Monticello, the Virginia home where Thomas Jefferson's dream of the New World as an Arcadian agricultural idyll fell apart during the "Tambora period".

Also making an appearance is John Barrow. This scheming second secretary to the British Admiralty successfully petitioned his employers to resume the quest for a North West Passage on the basis of whaler William Scoresby's reports of ice-free waters off northern Canada in 1817, with support from naturalist Joseph Banks at the Royal Society. With Napoleon defeated, Navy men needed to be kept busy. Many, Rear Admiral John Franklin being the best known, were sent to their deaths in the following decades: Tambora's far-flung climatic effects having waned, the ice returned with a vengeance and destroyed them.

This book is much more than brilliant popular science. Drawing together a world of data relating to this epoch-changing eruption, Wood has made a major contribution to volcanology, climatology and cultural history, driven by personal passion and conviction. ■

Ted Nield *edits the magazine* Geoscientist *for the Geological Society of London. His book* Underlands *(Granta Books) is due out in May.*
e-mail: ted.nield@geolsoc.org.uk



Semipalmated sandpipers (*Calidris pusilla*) on their annual autumn migration.

YVA MOMATIUK & JOHN EASTCOTT/MINDEN PICTURES/FLPA

ANIMAL BEHAVIOUR

Nomads of necessity

Joel Greenberg casts an ornithologist's eye on a wide-ranging reading of animal migration.

Animal migration is a wondrous thing. Driven by seasonal changes in daylight and other factors, an incredible array of organisms fly, creep or run through urban and other habitat-restricted areas on their way to or from richer spots for wintering, feeding or breeding. When you watch the movement of migrants (and I know birds best), you see the beauty of the individuals, and are privy to the performance of ancient commands dictated by genetic coding. It is often a species' most vulnerable moment. Human-driven and other shifts can occur too quickly for an individual to adjust: a bird that has flown all night looks down at the landscape that emerges at dawn, and seeks an accommodating wood that may or may not still be standing.

In *The Homing Instinct*, biologist Bernd Heinrich explores the concept of 'home' — how organisms in a wide range of taxa, including insects, arachnids, amphibians, fish, birds and mammals, arrive at their destinations, construct or alter home territories, and live in them. He discusses the placement of nests according to geographic and site-specific factors; nest structure and hygiene; parasitism and feeding behaviour. In the telling, he relies on the latest scientific advances as well as his own field observations.



The Homing Instinct: Meaning and Mystery in Animal Migration
BERND HEINRICH
Houghton Mifflin Harcourt: 2014.

Despite the tens of thousands of human eyes trained on animal migrants, many elements of the migration period, such as routes and duration of travel, were difficult to investigate until the advent of satellite technology. This has shed light on astounding behaviours, and Heinrich discusses one of my favourite cases: the prodigious speed and endurance of the bar-tailed godwit (*Limosa lapponica*).

This long-billed, long-legged shorebird was known to nest in the Arctic and winter in Australia and New Zealand. In a 2009 study, transmitters were placed on 23 of the birds to track their flight paths. They revealed that the species can fly from Alaska to New Zealand in less than nine days without stopping, riding wind currents and expending half their body weight (R. E. Gill Jr *et al. Proc. R. Soc. B* **276,** 447–457; 2009). (On the way north, the birds do stop to feed.)

Heinrich discusses at length how ▶

© 2014 Macmillan Publishers Limited. All rights reserved

species such as sandhill cranes (*Grus canadensis*) and bees return to the places of their birth. But from an evolutionary standpoint, it is important for some individuals not to get homing quite right.

Take salmon. These fish use smell to find their streams of origin, as was corroborated by experiments in which they were raised in hatcheries and exposed to a particular chemical, then released into Lake Michigan. Each of their spawning streams was treated with one of the chemicals, and indeed 90% of the salmon selected the stream with the smell of home. It is likely that most of the remainder did not survive; possibly they selected the 'correct' waterway, but it had become unsuitable for spawning. However, some of these mostly suicidal explorer fish would have survived, by finding different, habitable streams. This is essential for the long-term survival of the species.

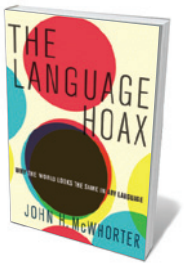
I think Heinrich could have left out the chapter 'Homing to the Herd', in which he switches his discussions from homing based on place to homing based on the flock. He sees the mass gatherings of the now-extinct Rocky Mountain locusts (*Melanoplus spretus*) as the home, even though they shifted from place to place. And in his treatment of the also extinct passenger pigeon (*Ectopistes migratorius*), which up to its demise in the late 1870s gathered in everything from pairs to aggregations of billions, he gets some important facts wrong.

It is true that passenger pigeons "were reputed to nest for most of the year wherever and whenever they found food". But those so reputing were often hunters or dealers in the game-meat industry, who wanted to allay fears that the birds were disappearing. In any event, as the naturalist A. W. Schorger (whom Heinrich cites as a reference) made clear, passenger pigeons generally nested only once a year. Heinrich bases much of this discussion on a comparison with the mourning dove (*Zenaida macroura*); however, genetic work has shown the passenger pigeon to be much closer to a group that includes the band-tailed pigeon (*Patagioenas fasciata*).

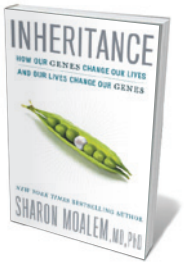
Having said all that, this is an informative and fascinating book. In choosing to explore this elemental behaviour across such a vast range of taxa, Heinrich has written a work for which many will provide a home. ■

Joel Greenberg *is research associate at the Peggy Notebaert Nature Museum and the Field Museum in Chicago, Illinois. He is the author of* A Feathered River Across the Sky: The Passenger Pigeon's Flight to Extinction.
e-mail: joelrgreenberg@gmail.com

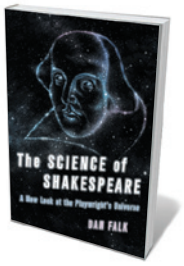
Books in brief



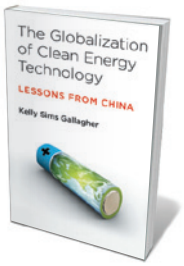
The Language Hoax: Why the World Looks the Same in Any Language
John H. McWhorter OXFORD UNIVERSITY PRESS *(2014)*
In this manifesto for a logical linguistics, one of John H. McWhorter's many intriguing examples hinges on the colour blue. To an English speaker, it is one subtly graded hue, from robin's-egg to deep navy. To a Russian, pale and dark blue are different colours. But does this point to crucial differences in how Russian and English speakers perceive? Emphatically not, avers McWhorter. He tackles such linguistic determinism — the Sapir–Whorf hypothesis — head on, arguing that world views are human, not strapped to one culture.



Inheritance: How Our Genes Change Our Lives — and Our Lives Change Our Genes
Sharon Moalem GRAND CENTRAL PUBLISHING *(2014)*
Physician Sharon Moalem delivers heavyweight genetics with a light touch in this study of "flexible inheritance" — the perpetual tango between genes and experience. He uses diagnostic readings of his patients (a bluish white of the eye, for instance, can indicate the bone-weakening condition osteogenesis imperfecta) as springboards for broader discussion on the complexities of genetic expression, 'handedness' and more. The idea of the genetic life sentence is shifting, he argues, even with some serious disorders.



The Science of Shakespeare: A New Look at the Playwright's Universe
Dan Falk THOMAS DUNNE BOOKS *(2014)*
As the 450th anniversary of Shakespeare's birth looms, many are debating whether the bard's plays reflect the new science of his era — a point that historian Jennifer Rampling has made in these pages (*Nature* **508,** 39–40; 2014). Here, science writer Dan Falk weighs in, combing the dramas for traces of findings such as heliocentricity. Falk speculates entertainingly as he confers with scientists and Shakespeare experts, some sceptical, some gung-ho about the possibilities. But as with so much regarding the enigma from Stratford-upon-Avon, Falk's conclusion is ambiguous.



The Globalization of Clean Energy Technology: Lessons from China
Kelly Sims Gallagher MIT PRESS *(2014)*
The need for clean-energy innovation has never been greater, yet pundits argue that trade barriers and intellectual-property issues hinder technology diffusion. Kelly Sims Gallagher's study of China — the world's leading energy consumer, and a hotbed for the production and export of imported technologies — offers nuance. Her take on the country's solar photovoltaic industry, for instance, busts myths about barriers, revealing that channels such as joint research and development have globalized technology transfer. The real issue, she argues, is a lack of viable government-policy incentives.



Are We All Scientific Experts Now?
Harry Collins POLITY PRESS *(2014)*
Forty years of work with gravitational-wave physicists has given sociologist Harry Collins an unusual perspective on scientific expertise, as chronicled in *Gravity's Ghost and Big Dog* (University of Chicago Press, 2013; see M. de Laet *Nature* **501,** 164–165; 2013). Here, he distils his findings to clarify why specialized scientific knowledge is essential for all. Partly a polemic for lay readers still aghast at Climategate, this is also a reminder of why the values and aspirations of lab scientists matter. **Barbara Kiser**

© 2014 Macmillan Publishers Limited. All rights reserved

Correspondence

China leads the way on renewables

The boom in renewable energy being led by China, India, Europe, the United States and Japan is key to the mitigation of carbon emissions (see *Nature* **507,** 300–302; 2014). Last year marked an important turning point for China's renewable-energy revolution in particular, with water, wind and solar sources accounting for more electricity-generating capacity (59%) than new fossil-fuel or nuclear facilities (41%; see go.nature.com/z6job5).

China generated 5,322 billion kilowatt-hours of electricity, of which 74% came from coal-fired power stations (a marked improvement on the 80% mentioned by some researchers; see D. Helm *Nature* **491,** 663–665; 2012). The balance was contributed primarily by hydropower (17%), wind (2.6%), nuclear (2.1%) and solar sources (0.16%). Water, wind and solar energy therefore account for one-fifth of electricity generated in China — anticipating by two years the target that the country set for 2015.

John A. Mathews *Macquarie University, Sydney, Australia.*
john.mathews@mgsm.edu.au
Hao Tan *University of Newcastle, Callaghan, New South Wales, Australia.*

Vanadium batteries will be cost-effective

Vanadium flow batteries are an attractive commercial proposition because they are safe and environmentally friendly, use recyclable electrolytes, have a long cycle life (around 13,000 cycles) and last for more than 15 years. Their cost is not as prohibitive as believed (*Nature* **507,** 26–28; 2014).

You quote my estimated cost of US$1,000 per kilowatt-hour, which is indeed high, but this applies to the world's largest vanadium flow battery system and includes materials, manufacture, the control system, transportation, installation and maintenance. The cost of a system that is one-fifth of this size is $600 per kilowatt-hour, which is expected to drop to $400 per kilowatt-hour within the next 2–3 years.

After two years of innovation and development, the current density of vanadium flow battery stacks from the Rongke Power company in Dalian, China, has risen from 80 to 120 milliamps per square centimetre.

Although the cost of vanadium itself is relatively high, these flow batteries hold promise as large-scale energy-storage devices.

Huamin Zhang *Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China.*
zhanghm@dicp.ac.cn

A trading scheme to reduce false results

Cap-and-trade systems have proved useful in cutting pollutants such as sulphur dioxide, nitrogen oxides and lead additives in petrol (see L. H. Goulder *J. Econ. Perspect.* **27,** 87–102; 2013). We suggest that they could also be applied to reduce pollution of the scientific literature with irreproducible results (see go.nature.com/huhbyr).

Companies do not have to deal with the social costs of pollution, so there is little incentive for them to reduce it. Likewise, researchers do not have to face the cost of publishing their own unverifiable results (most of which could have been prevented). That cost is borne by the scientific community and the public — for instance, in subsequent research inspired by false positives, which can lead to badly designed policies.

Cap-and-trade systems force excessive polluters to purchase permits. Initially, institutions could receive 5 free permits per 100 published results, reflecting the widely accepted ideal of a 5% false-positive production rate. It would then be necessary to buy extra permits from other institutions should they 'emit' significantly more false positives that this (irrespective of whether these were honest or deliberate errors).

Institutions that successfully reduce false positives in their research output could then sell off their surplus permits to other institutions that have exceeded their allocation. This flexibility would create incentives for researchers to find innovative ways to reduce false positives.

Michael E. McCullough, David L. Kelly *University of Miami, Coral Gables, Florida, USA.*
mikem@miami.edu

Mouse already being charted gene by gene

Steve Perrin's call to make mouse studies work (*Nature* **507,** 423–425; 2014) resonates with the goals and practices of the International Mouse Phenotyping Consortium (see www.impc.org).

The consortium's ten-year goal is to generate a 'knockout' mutant for every gene in the mouse genome in an effort to characterize the phenotype that each gene confers (S. D. Brown and M. W. Moore *Mamm. Genome* **23,** 632–640; 2012). Standardized phenotyping protocols (www.impc.org/impress) have been carefully designed and validated by the consortium to provide robust, reproducible information, and a statistics advisory group reviews the data and procedures.

All mutant lines and phenotype data are freely available, including summary data for a cohort compared with multiple controls. To overcome any potential issue of publication bias, we include all negative results as well as positive ones.

Careful evaluation of mouse models and their standardized phenotypes, as well as confidence in the reproducibility and validity of summary and individual phenotypic data, is critical to fostering successful preclinical studies.

Steve Brown* *International Mouse Phenotyping Consortium; MRC Harwell, Didcot, UK.*
s.brown@har.mrc.ac.uk
**On behalf of 22 co-signatories; see go.nature.com/iu6xuf for full list.*

Experimentation is science's lifeblood

No one can become a genuine scientist without doing practical work, as John Baruch points out (see *Nature* **507,** 141; 2014) — whether it is at the lab bench or desk. But it is not the successful experiments that count so much as the frustrations inflicted by those that fail.

It is the embarrassment of a statistical analysis that reveals inadvertent data-input errors; it is tearing one's hair out debugging a program that once worked well; it is learning that "nature cannot be fooled" (as physicist Richard Feynman warned).

This experience of failure and error is unique to science. In the humanities, a student learns from mistakes such as not remembering to use specialist vocabulary, failing to put forward established arguments or not engaging in self-promotion before and after seminars. Practical science demonstrates that doing science involves more than just picking up such tricks.

John Skoyles *University College London, UK.*
j.skoyles@ucl.ac.uk

CONTRIBUTIONS

Correspondence may be sent to **correspondence@nature.com** after consulting the author guidelines at **http://go.nature.com/cmchno**.

© 2014 Macmillan Publishers Limited. All rights reserved

EVOLUTIONARY BIOLOGY

Dating chimpanzees

Genetic research has tracked lineages of male chimpanzees thousands of years into the past, opening the door to the study of long-term behavioural evolution in our close primate relatives.

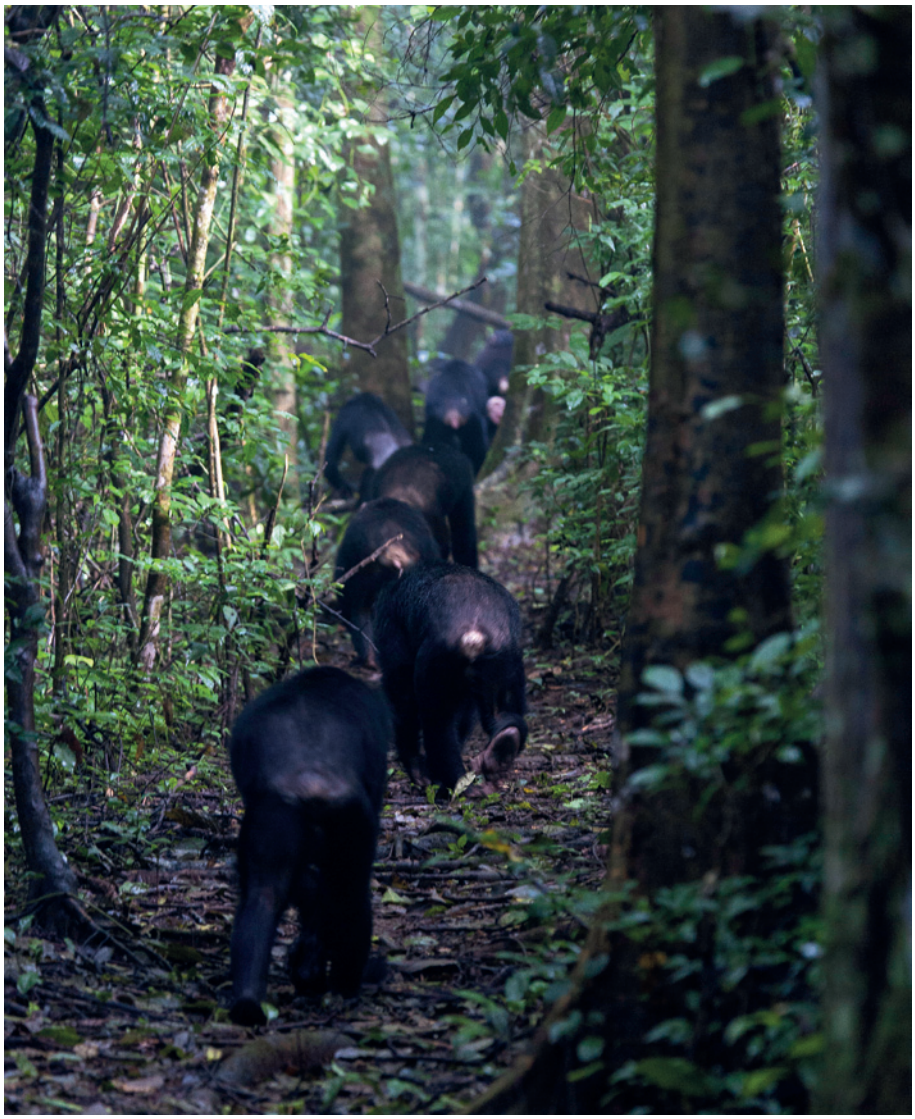
MICHAEL HASLAM

To a casual observer, one chimpanzee can seem much the same as another. Yet there are four distinct subspecies of chimpanzee living across Africa today, each of which is made up of many discrete communities. Within these groups, specific behavioural characteristics are shared from one generation to the next, resulting in a patchwork of behaviours that bears the hallmarks of cultural transmission[1]. Recent studies have shown how such transmission may occur, in both experimental[2] and wild[3] groups. But one crucial factor was missing: time. Now, writing in the *Journal of Human Evolution*, Langergraber *et al.*[4] show that some modern chimpanzee communities have probably existed for hundreds or even thousands of years. By providing a chronological scaffold for chimpanzee behaviour, this research offers a rare opportunity to examine the tempo of chimpanzee cultural evolution.

Langergraber and colleagues used genetic markers on the Y chromosome, which are passed only from father to son, to examine the relationships between eight communities of East African chimpanzees (*Pan troglodytes schweinfurthii*) in Uganda (Fig. 1). Chimpanzees are male philopatric, which means that males tend to stay in the territory in which they were born and adolescent females emigrate to a new community. The stable core of a group is therefore composed of related males — an unusual pattern among primates, and a potential contributor to high levels of male aggression between chimpanzee groups[5]. Y chromosomes accumulate mutations over time, resulting in the build-up of community-specific shared mutations known as haplotypes. Combined with an estimate of the rate of mutation, haplotype variations allow calculation of the time that has elapsed since any pair of individuals shared a common ancestor.

Large-scale genetic comparisons among the great apes have suggested that chimpanzees probably split from bonobos around 2 million years ago[6]. Concentrating on much finer timescales, Langergraber *et al.* estimate a time to the most recent common ancestor (TMRCA) ranging between 125 and 2,625 years for their eight chimpanzee communities. Because chimpanzee groups are likely to be founded by small numbers of individuals splitting from larger communities, the authors argue that the TMRCA estimates are close to the time that those groups were established. They even uncovered evidence for a likely fission event: two neighbouring communities, the Ngogo and Kanyantale groups in Kibale National Park, share a haplotype and a TMRCA of about 450–500 years ago, suggesting that these two groups may have split from one another around this time. Future work will be necessary to refine the wide confidence intervals surrounding these estimates.

The findings clearly show that East African chimpanzee groups display great stability and longevity. For example, the founding of the



FLORIAN MOELLERS/RZSS

Figure 1 | Stable communities. The Sonso group of East African chimpanzees (*Pan troglodytes schweinfurthii*), from the Budongo Forest in Uganda, is estimated to be the oldest of the eight chimpanzee communities studied by Langergraber and colleagues[4].

© 2014 Macmillan Publishers Limited. All rights reserved

oldest sampled community, the Sonso group in the Budongo Forest, at more than 2,500 years ago, may pre-date the Roman Empire, the Classic Maya and the Chinese Han Dynasty. On this evidence, the Sonso chimpanzees could be considered to have one of the world's oldest continuous cultures. However, the techniques and results presented by Langergraber *et al.* offer significant potential beyond simply estimating the time of community-founding events.

Chimpanzee cultural differences are maintained through social learning, in which group members conform to observed behaviour, either during their juvenile development[3] or as newly immigrant females[7]. Although females act as an important cultural vector, bringing new behaviours into a group when they arrive, it is possible that the resident males act as a brake on the speed at which cultural variation accrues. Without such a conservative mechanism, cultural differences between groups would be eroded with each passing generation, and the pattern we observe today would not have emerged. A similar mechanism was probably in place among early human ancestors, if stasis in the development of flaked-stone technology over its first million years is a guide[8].

Hypothetically, in the absence of a conservative tendency, longer-lived groups should accrue more cultural idiosyncrasies. But the new genetic data cast doubt on that hypothesis. For example, the Kanyawara and Ngogo groups in Kibale National Park are located less than 20 kilometres apart, and have similar numbers of observed cultural variants, yet the TMRCA estimates are more than 1,700 years for the former and close to 450 years for the latter. Once similar genetic data are available for other sites and subspecies, it will be possible to assess more clearly whether there is a pattern to how quickly a new community reaches a certain level of cultural complexity. We may also be able to estimate age ranges for the origin of behaviours shared by communities in a specific region, such as stone-tool use by West African chimpanzees[9].

The availability of dated communities also allows us to look more closely at the reasons for their founding. For example, fluctuation in African forest sizes through time probably influences chimpanzee community dynamics[10]. Keeping error margins in mind, it is interesting that for the five chimpanzee groups with a TMRCA of less than 1,000 years, these dates all coincide with periods of East African drought[11]. Do droughts promote chimpanzee group fission, as forests fragment and food scarcity increases? The same factors might also negatively affect the retention of cultural knowledge related to the exploitation of particular insect or nut species, creating a complex temporal link between genetic and cultural evolution.

Genetic dates such as those provided by Langergraber *et al.* open up a complementary perspective to that provided by the past 50 years of field observations of wild chimpanzees. They also demonstrate that the array of approaches used to reconstruct human evolution is applicable to chimpanzees, aiding the study of primate archaeology[12]. Clearly, the process of exploring the fine-grained history of non-human animal behaviour has only just begun. ■

Michael Haslam *is in the Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK.*
e-mail: michael.haslam@rlaha.ox.ac.uk

1. Whiten, A. *et al. Nature* **399,** 682–685 (1999).
2. Whiten, A., Horner, V. & de Waal, F. B. M. *Nature* **437,** 737–740 (2005).
3. Biro, D. *et al. Anim. Cogn.* **6,** 213–223 (2003).
4. Langergraber, K. E. *et al. J. Hum. Evol.* http://dx.doi.org/10.1016/j.jhevol.2013.12.005 (2014).
5. Manson, J. H. *et al. Curr. Anthropol.* **32,** 369–390 (1991).
6. Prüfer, K. *et al. Nature* **486,** 527–531 (2012).
7. Luncz, L. V. & Boesch, C. *Am. J. Primatol.* http://dx.doi.org/10.1002/ajp.22259 (2014).
8. Stout, D., Semaw, S., Rogers, M. J. & Cauche, D. *J. Hum. Evol.* **58,** 474–491 (2010).
9. Haslam, M. *Antiquity* **86,** 299–315 (2012).
10. Gruber, T. *et al. J. Comp. Psychol.* **126,** 446–457 (2012).
11. Verschuren, D., Laird, K. R. & Cumming, B. *Nature* **403,** 410–414 (2000).
12. Haslam, M. *et al. Nature* **460,** 339–344 (2009).

GENETICS

Up and down in Down's syndrome

A comparison of identical human twins, only one of whom has Down's syndrome, reveals a genome-wide flattening of gene-expression levels in the affected individual. SEE ARTICLE P.345

BENJAMIN D. POPE & DAVID M. GILBERT

Down's syndrome occurs when humans have an extra copy of chromosome 21 (ref. 1), a situation referred to as trisomy 21. Because each chromosome contains a distinct set of genes that serve as blueprints for the expression of cellular components, it has been presumed for decades that the condition is mainly caused by an overabundance of the products of chromosome 21 genes. But on page 345 of this issue, Letourneau *et al.*[2] report a case of Down's syndrome that is associated with altered gene expression across every chromosome, not just chromosome 21. This observation implies that the expression of any number of genes on any chromosome may contribute to Down's syndrome, and raises the possibility that an extra copy of any chromosome can disrupt general gene regulation.

The discovery was made possible by an elegantly controlled experiment that compared a set of twins derived from the same fertilized egg (monozygotic, or 'identical', twins) in which one twin had an extra chromosome 21 and the other did not, owing to chromosome-segregation errors that occurred before the twinning event[3]. This unusual circumstance allowed the effects of the extra chromosome 21 to be studied in isolation. Although gene expression has been extensively studied in individuals with Down's syndrome, the genome-wide effect discovered by Letourneau *et al.* had gone undetected because, as they show, natural variation among individuals

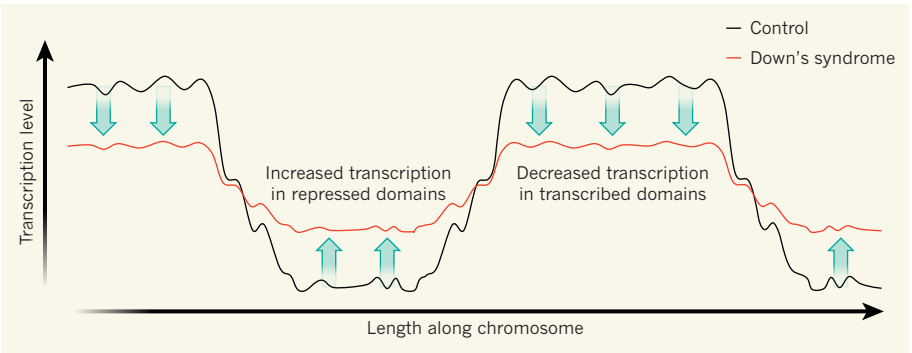


Figure 1 | Flattened gene expression. Letourneau *et al.*[2] show that genomic domains that are normally associated with low or high levels of gene expression are respectively up- or downregulated in a person with Down's syndrome, compared with their identical twin who does not have the condition. The result is a genome-wide flattening of gene expression.

© 2014 Macmillan Publishers Limited. All rights reserved

is strong enough to mask the effect.

Importantly, the authors found that the altered gene expression followed a consistent pattern, with increased and decreased gene-expression levels alternating across large chromosomal segments. The discovery of these up- and downregulated segments, which Letourneau *et al.* call gene expression dysregulation domains (GEDDs), supports mounting evidence that chromosomes contain functional domains that may help to provide cells with access to the genetic information at the appropriate place and time. The positions of the GEDDs align with chromosome domains defined by other structural and functional properties, such as domains that associate with nuclear lamina proteins (lamina-associated domains; LADs[4]) or that are replicated at different times during the DNA-synthesis phase of the cell-division cycle[5]. These findings strengthen the idea that chromosome functions reflect underlying structural domains.

The authors also report the presence of GEDDs in mice that carry an extra piece of chromosome 16 (the mouse counterpart to most of human chromosome 21) and that show several features of Down's syndrome[6]. The GEDDs were observed throughout the mouse genome at positions corresponding to their locations on human chromosomes. Furthermore, the authors demonstrate that the domains were largely preserved after the human twins' cells were artificially reprogrammed to a developmental state resembling that of embryonic stem cells (induced pluripotent stem cells)[7]. The authors understandably focus on the similarities between GEDDs before and after this reprogramming. However, the differences that they observed may correspond to the changes in replication timing, or perhaps lamina association, that occur during reprogramming[5].

Intriguingly, Letourneau and colleagues show that GEDDs with increased expression corresponded to otherwise repressed genomic domains, whereas GEDDs with decreased expression corresponded to domains normally characterized by active transcription (Fig. 1). This means that there is a diminished difference between expressed and repressed genes in people with Down's syndrome, suggesting that the extra chromosome 21 interferes with the cell's ability to regulate transcriptional output.

The authors made several attempts to understand the mechanism behind GEDDs, but they found no significant changes in LADs or in patterns of DNA methylation — a modification that affects gene-transcription rates. They did find that levels of trimethylation at amino-acid residue lysine 4 on histone H3 correlated well with the transcriptional changes seen in GEDDs (histones are proteins around which DNA is wound in the nucleus, forming a complex called chromatin), but this is to be expected because such post-translational histone modification tracks with expressed genes[8]. The results of the authors' investigation of chromatin accessibility within GEDDs (the accessibility of chromatin to gene-transcription machinery also regulates expression levels) were difficult to interpret.

So how could the addition of a single, relatively small chromosome — chromosome 21 is the smallest human chromosome and accounts for less than 2% of the genome — dampen transcriptional differences across the genome? Two kinds of mechanism seem most plausible. First, and perhaps most simply, it is possible that the increased dosage of one or more genes on chromosome 21 is responsible. For example, human chromosome 21 and mouse chromosome 16 carry the *HMGN1* gene, the product of which competes[9] with histone H1 for access to the linker DNA between nucleosomes, the repeating units of chromatin. Because H1 is associated with less-accessible chromatin, an increase in dosage of *HMGN1* would be consistent with an increase in global chromatin accessibility. Increased access to normally inaccessible chromatin would be expected to dilute the activity of factors that switch on genes in other parts of the genome, or release factors that repress genes in active regions, or both, with the net effect of flattening gene-expression levels genome-wide. An obvious experiment would be to examine the effect of controlled overexpression of *HMGN1* on global transcription levels.

A second, much less defined possibility is that the phenomenon described by Letourneau *et al.* results from the extra DNA content, for example by sequestering factors that regulate expression[10]. This hypothetical mechanism need not be specific to chromosome 21 and could be explored further by comparing monozygotic twins that differ in other trisomies. Although less common, most other trisomies do cause some of the clinical features of Down's syndrome, and extra copies of larger chromosomes are associated with more-extreme effects[11]. Sex chromosomes, which are much more benign in a trisomic context, are the exception[11].

Letourneau and colleagues have used a set of well-controlled, carefully performed and reproducible experiments to report a provocative new phenomenon. Their findings raise more questions than they answer, and open the door to exciting further research. ■

Benjamin D. Pope *and* **David M. Gilbert**
are in the Department of Biological Science, Florida State University, Tallahassee, Florida 32306, USA.
e-mails: pope@bio.fsu.edu; gilbert@bio.fsu.edu

1. LeJeune, J., Gautier, M. & Turpin, R. *C.R. Hebd. Séanc. Acad. Sci.* **248,** 602–603 (1959).
2. Letourneau, A. *et al. Nature* **508,** 345–350 (2014).
3. Dahoun, S. *et al. Am. J. Med. Genet. A* **146A,** 2086–2093 (2008).
4. Guelen, L. *et al. Nature* **453,** 948–951 (2008).
5. Hiratani, I. *et al. PLoS Biol.* **6,** e245 (2008).
6. Davisson, M. T. *et al. Prog. Clin. Biol. Res.* **384,** 117–133 (1993).
7. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
8. Li, B., Carey, M. & Workman, J. L. *Cell* **128,** 707–719 (2007).
9. Catez, F., Brown, D. T., Misteli, T. & Bustin, M. *EMBO Rep.* **3,** 760–766 (2002).
10. Liu, X., Wu, B., Szary J., Kofoed, E. M. & Schaufele, F. *J. Biol. Chem.* **282,** 20868–20876 (2007).
11. Hassold, T. & Hunt, P. *Nature Rev. Genet.* **2,** 280–291 (2001).

ORGANIC CHEMISTRY

Catalysis marches on

A fresh take on an established chemical reaction has solved a long-standing problem in organic synthesis: how to prepare single mirror-image isomers of groups known as isolated quaternary stereocentres. SEE ARTICLE P.340

JAMES P. MORKEN

The production of a wide array of compounds, ranging from polymers to liquid crystals to human therapeutics, depends on stereoselective synthesis, which enables three-dimensional control over the isomer of the product that forms. Not surprisingly, the more stereoselective reactions that chemists have in their toolbox, the more efficiently they can construct these materials. One tool that is not well developed is a method for the stereoselective construction of quaternary stereocentres — in which a carbon atom is bonded to the carbon atoms of four other distinct appendages — at positions that are remote from any other groups in a molecule. In this issue, Sigman and colleagues[1] (page 340) describe a process that accomplishes just that.

The handedness, or chirality, of enzymes, proteins, nucleic acids and carbohydrates results in distinct binding sites that often accommodate one mirror-image form (enantiomer) of a small-molecule substrate better than the other. Preparing effective inhibitors of biological processes is therefore dependent on our ability to selectively make one enantiomeric form of a small molecule. Such selectivity is the goal of asymmetric organic synthesis. A central dogma in this area is that optimum efficiency results from catalytic asymmetric reactions, in which the handedness of the catalyst

© 2014 Macmillan Publishers Limited. All rights reserved

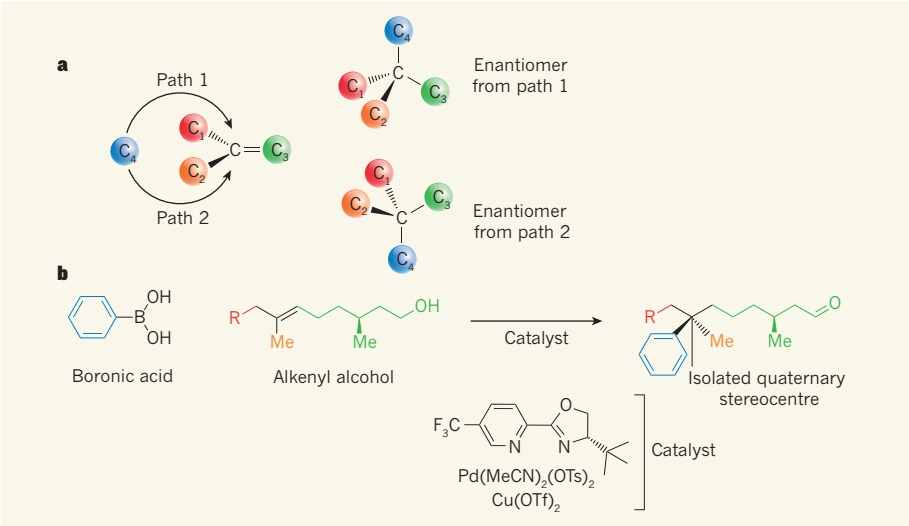


Figure 1 | Enantiomeric synthesis of quaternary stereocentres. Many organic compounds can be arranged in mirror-image forms called enantiomers. **a**, The two enantiomers of chiral quaternary stereocentres are produced by different pathways that connect a fourth carbon appendage (C_4) to a carbon atom bearing three others (C_1, C_2 and C_3). This requires a catalyst (not shown), which must be reactive enough to overcome energy barriers to create the congested centre, and do so preferentially by one path. **b**, Sigman and colleagues[1] report a palladium catalyst that enables the reaction of boronic acids with alkenyl alcohols to form quaternary stereocentres that are located remotely from any other reactive groups. Me, methyl; R, methyl, ethyl, propyl or $OSiMe_2CMe_3$, an alcohol derivative; OTs, tosylate ($CH_3C_6H_4SO_3^-$); OTf, triflate ($CF_3SO_3^-$).

controls the handedness of the reaction product. Asymmetric catalysis thus allows small amounts of single-enantiomer catalysts to produce large amounts of single-enantiomer products.

Even after enormous effort and significant advances in the field of chemical catalysis, there are still many more problems than there are solutions. One large obstacle is the enantioselective construction of quaternary stereocentres[2]. The challenge in making these features is two-fold. First, the fourth and final carbon appendage must be attached to a 'hindered' carbon centre that has three other carbon appendages arranged around it (Fig. 1a). Such carbon–carbon (C–C) bond-forming reactions face substantial energy barriers that can defy even the most reactive of reagents. The second challenge is to complete the construction in a way that gives just one enantiomer of the product. This requires the catalyst to distinguish the size, shape and properties of the three pre-existing appendages, and to do so with high fidelity.

A common reaction design for constructing chiral carbon centres involves adding carbon atoms to one end of a carbon–carbon double bond in an alkene molecule, using a catalyst both to facilitate bond formation and, by distinguishing between the groups attached to the double bond, to control the handedness of the product. Such discrimination is especially challenging in the synthesis of quaternary centres, however, because the groups to be distinguished are all carbon substituents and are therefore similar in size.

To accelerate such reactions, chemical groups are often attached to the reacting alkene, a manoeuvre that can also help the catalyst to discriminate between the two faces of planar alkene groups, and so render reactions more enantioselective. For example, a carbonyl group (C=O) can be attached to the alkene to enable conjugate addition, a widely used process for forming C–C bonds[3,4]. Alternatively, a carbon bearing a leaving group (a group that departs from a molecule as an anion during a reaction) can be attached, thereby allowing a different kind of C–C bond-forming reaction, the allylic substitution reaction[5,6], which is a topic of research in my lab[7]. A consequence of this group-attachment manoeuvre, however, is that the newly formed C–C bond necessarily bears a neighbouring group, which might not be desired.

Sigman and colleagues now describe a palladium catalyst that facilitates C–C bond formation between trisubstituted alkenyl alcohols and boronic acids under oxidative conditions (Fig. 1b), and which leads to products that do not have appended neighbouring groups. The process is based on a known variant[8] of a powerful C–C bond-forming process catalysed by a transition metal, which was developed by Tsutomu Mizoroki[9] and Richard Heck. (Heck won a share of the 2010 Nobel Prize in Chemistry for work in this area[10].) Sigman and colleagues describe the first such reaction that allows C–C bond formation to generate isolated quaternary centres intermolecularly and with outstanding levels of enantiomeric selectivity.

Concomitant with C–C bond formation, the double bond in the alkenyl alcohol migrates by 'walking' along the molecule's carbon chain, until it encounters an alcohol group (OH) farther down the molecule and converts it to a carbonyl group. A powerful feature of the reaction is that the chain walking can occur over many carbon atoms, and allows one to choose the spatial relationship between the newly formed quaternary centre and the resulting carbonyl. Such long-range chain-walking events have previously been observed in polymerization reactions[11], but have not been widely used as a design element in organic synthesis[12].

One question that Sigman and co-workers address is what happens when the catalyst, engaged in the act of chain walking, encounters a pre-existing chiral carbon centre? If the catalyst releases the alkene as chain walking occurs, then undesirable scrambling of the carbon centre to form an equal mixture of enantiomers is guaranteed. The authors find that the configuration of such chiral centres is retained with perfect selectivity, an outcome that suggests that tight binding occurs between the alkene and palladium as the catalyst speeds down the hydrocarbon backbone towards the carbonyl.

Sigman and colleagues' report sows the seeds of a new direction in asymmetric Mizoroki–Heck reactions, but there is clearly much to do for this area to grow. Currently, the reaction allows only aryl groups — benzene rings with or without other groups attached — to attach to the alkenyl alcohol. Expanding the scope of the reaction to allow the attachment of other groups, such as saturated and unsaturated hydrocarbons, will further extend its utility. It will also be interesting to learn whether groups other than alcohols can intercept the chain-walking palladium catalyst, because this might allow the reaction to be diverted in new and useful directions. Nevertheless, construction of quaternary stereocentres just became much easier. ■

James P. Morken *is in the Department of Chemistry, Boston College, Chestnut Hill, Massachusetts 02467, USA.*
e-mail: morken@bc.edu

1. Mei, T.-S., Patel, H. H. & Sigman, M. S. *Nature* **508,** 340–344 (2014).
2. Das, J. P. & Marek, I. *Chem. Commun.* **47,** 4593–4623 (2011).
3. Alexakis, A., Bäckvall, J. E., Krause, N., Pàmies, O. & Diéguez, M. *Chem. Rev.* **108,** 2796–2823 (2008).
4. Harutyunyan, S. R., den Hartog, T., Geurts, K., Minnaard, A. J. & Feringa, B. L. *Chem. Rev.* **108,** 2824–2852 (2008).
5. Hoveyda, A. H., Hird, A. W. & Kacprzynski, M. A. *Chem. Commun.* 1779–1785 (2004).
6. Lu, Z. & Ma, S. *Angew. Chem. Int. Edn* **47,** 258–297 (2007).
7. Zhang, P., Le, H., Kyne, R. E. & Morken, J. P. *J. Am. Chem. Soc.* **133,** 9716–9719 (2011).
8. Du, X. *et al. Org. Lett.* **3,** 3313–3316 (2001).
9. Mizoroki, T., Mori, K. & Ozaki, A. *Bull. Chem. Soc. Japan* **44,** 581 (1971).
10. Heck, R. F. & Nolley, J. P. Jr *J. Org. Chem.* **37,** 2320–2322 (1972).
11. Johnson, L. K., Killian, C. M. & Brookhart, M. *J. Am. Chem. Soc.* **117,** 6414–6415 (1995).
12. Kochi, T., Hamasaki, T., Aoyama, Y., Kawasaki, J. & Kakiuchi, F. *J. Am. Chem. Soc.* **134,** 16544–16547 (2012).

This article was published online on 9 April 2014.

© 2014 Macmillan Publishers Limited. All rights reserved

SYNTHETIC BIOLOGY

Biocircuits in synchrony

Cellular biocircuit design has taken a major step forward. The circuit reuses the cell's own protein-degradation system to synchronize the expression of two synthetic modules throughout an entire bacterial population. SEE LETTER P.387

RICARD SOLÉ & JAVIER MACÍA

A major goal of synthetic biology is to build reliable, predictable networks of molecular and cellular components that can work as new biological devices capable of, for example, sensing chemicals, manufacturing drugs or even fighting disease. However, achieving such goals entails the production of complex synthetic biocircuits, which requires synchronization of multiple components. Although synchronization is well established in electronics[1], synchronizing living cells is a major challenge, because it demands correlation of different phenomena that may be taking place on different temporal and spatial scales. On page 387 of this issue, Prindle *et al.*[2] report that such coupling has been achieved in cells of the bacterium *Escherichia coli*.

In complex circuits, synchronization is required to equilibrate two or more modules that normally work on different timescales — if two components act at different speeds, synchronization ensures that the faster component waits until the slower one has finished its work before progressing, making it possible for the system to operate as a whole[3]. In electrical circuits, synchronization is achieved using an external 'clock signal'. In an electrical oscillator device, for example, when the clock signal is high (or low), the device stops oscillating, starting again only when the clock signal is low (or high). When applied to multiple components, the clock signal can therefore regulate the separate devices so that they work in unison.

This standardized approach to electrical circuits cannot, however, be directly exported to synthetic biocircuits, in which it is difficult to obtain a precise signal that equally affects all modules of the device, because the response of each module to an external stimulus is likely to vary. To solve this problem, Prindle and colleagues have developed an approach for fast and efficient synchronization of oscillators in *E. coli*, based not on an external clock, but instead using the existing machinery of the host cell as a synchronization system.

In general, oscillations arise from the periodic creation and destruction of a signal. In some synthetic bacterial oscillators, this signal is encoded in the concentration of the protein produced by the oscillator module — this protein typically inhibits the promoter that drives the module's own gene expression, a process known as negative feedback. Therefore, to produce oscillations, production of protein by the module must be continuously followed by the protein's degradation by a protease enzyme. Prindle and co-workers simultaneously introduced two previously reported synthetic oscillating modules into *E. coli*, which work on different temporal and spatial scales[4,5].

To produce oscillations at a single-cell level, Prindle *et al.* used an intracellular negative-feedback (NFB) module[5]. The NFB module produces λ-repressor protein, which negatively regulates the module, producing low-amplitude, high-frequency oscillations in each cell (Fig. 1a). The module then drives the expression of an output — in this case, a cyan fluorescent protein — resulting in the same low-amplitude, high-frequency oscillations.

To produce oscillations on a population-wide scale, the authors coupled the NFB module to a second oscillator that works through quorum sensing (QS)[6], the mechanism by which certain bacteria communicate. In QS, each bacterium secretes a signalling molecule into the extracellular medium, so that the concentration of the molecule in the medium depends on the local population density. Once the population reaches a threshold size, the concentration of the QS molecule increases to such a level that it is able to re-enter neighbouring cells, triggering gene transcription. The QS module used by Prindle and colleagues produces not only a QS signalling molecule, but also AiiA, a protein that acts as a repressor of the QS module, generating negative feedback. Expression of the QS module in a sufficiently large bacterial population results in simultaneous oscillation in each cell of the population at high amplitude and low frequency (Fig. 1b). This leads to production of the module's output — yellow fluorescent protein — in oscillations that follow the same pattern.

Prindle *et al.* achieved coupling between the two oscillators by exploiting a natural component of *E. coli*: the ClpXP protease enzyme[6]. The researchers avoided the need to engineer an extra regulatory component to couple the oscillators by tagging the fluorescent and feedback proteins produced by each module with a protein that allows degradation by ClpXP. Thus, protease activity regulates the level of signal available to feed back to each module, and the level of the module's output protein.

When the levels of protein produced by the QS module are low, enough ClpXP protease is available to degrade the λ-repressor protein, and the intracellular NFB oscillator fluctuates normally. But when the QS module is active, the total amount of protein generated by both oscillators saturates the degradation capacity of the ClpXP protease, and the intracellular

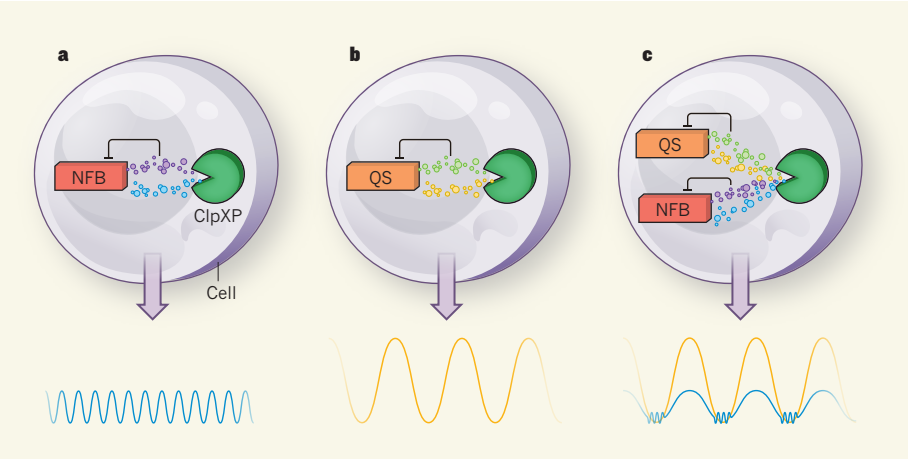


Figure 1 | Climbing up complexity in synthetic biology. **a**, The intracellular negative-feedback (NFB) oscillator is a synthetic module that works at a single-cell level in bacteria. The module produces λ-repressor protein (purple), which inhibits the oscillator's own activity, and a cyan fluorescent protein (blue). These proteins are degraded by the ClpXP protease enzyme, resulting in high-frequency, low-amplitude oscillations, measured by levels of cyan fluorescence (depicted by blue lines). **b**, A quorum-sensing (QS) module produces the negative feedback protein AiiA (green) and a yellow fluorescent protein (yellow), which are also degraded by ClpXP. Oscillation of the QS module is synchronized throughout the microbial population, and oscillations occur at low frequency and high amplitude (depicted by yellow lines). **c**, Prindle *et al.*[2] have generated the first complex synthetic biocircuit by simultaneously introducing these two oscillators into *Escherichia coli* cells. When both compete for degradation by ClpXP, their dynamics are synchronized.

© 2014 Macmillan Publishers Limited. All rights reserved

oscillator stalls (Fig. 1c). As a result, the activity of the intracellular module follows that of the QS module. Once the proteins produced by the QS module are degraded to lower levels, sufficient ClpXP is available to degrade λ-repressor protein again, and the intracellular oscillator returns to its normal, high-frequency oscillations.

Prindle and co-workers' approach does not require complex engineering to synchronize the oscillators because the authors have tinkered with the cell's own components. Moreover, whereas genetic devices are often designed to minimize the interactions between different components of the biocircuit, here these interactions have been strengthened. This strategy is appealing because the researchers' data suggest that exploration of the synergistic interactions between genetic devices and host cells will benefit circuit building in synthetic biology.

Prindle and colleagues' results provide an excellent argument for using the cell's natural machinery to integrate multiple synthetic components, because the authors have achieved fast, tunable and robust synchronization of two different modules. Complex decision-making circuits might strongly benefit from the authors' design. However, it is important to remember that scaling up the number of modules in a circuit remains a major issue within synthetic biology[7].

If circuit complexity is to grow in size and diversity, further improvements will be needed. For example, different strategies will be required to synchronize other types of genetic device, such as logic gates. The current work indicates that it may be possible to isolate different parts of a synchronized circuit in different cells, exploiting the cells as natural units of computation to perform basic logic functions such as AND or NOR. In this context, the use of different cell types communicating with one another to perform different functions might be a natural step forward[8–10], expanding on the method designed by Prindle *et al.* to allow generation of more complex, decision-making circuits. ■

Ricard Solé *and* **Javier Macía** *are in the Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona 08003, Spain.*
e-mails: ricard.sole@upf.edu; javier.macia@upf.edu

1. Purnick, P. E. & Weiss, R. *Nature Rev. Mol. Cell. Biol.* **10,** 410–422 (2009).
2. Prindle, A. *et al. Nature* **508,** 387–391 (2014).
3. Oklobdzija, V. G. *et al. Digital System Clocking: High-Performance and Low-Power Aspects* (Wiley, 2003).
4. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. *Nature* **463,** 326–330 (2010).
5. Stricker, J. *et al. Nature* **456,** 516–519 (2008).
6. Cookson, N. A. *et al. Mol. Syst. Biol.* **7,** 561 (2011).
7. Kwok, R. *Nature* **463,** 288–290 (2010).
8. Tamsir, A., Tabor, J. J. & Voigt, C. A. *Nature* **469,** 212–215 (2011).
9. Regot, S. *et al. Nature* **469,** 207–211 (2011).
10. Solé, R. V. & Macía, J. *Nat. Comput.* **12,** 485–497 (2013).

This article was published online on 9 April 2014.

THERMOELECTRICITY

The ugly duckling

Single crystals of tin selenide have been shown to display, along one crystallographic direction of their high-temperature state, the highest thermoelectric efficiency of any bulk material. SEE LETTER P.373

JOSEPH P. HEREMANS

More than 90% of the energy we use comes from thermal processes[1], which produce the bulk of the electricity generated by power plants, as well as powering aeroplanes and most cars. Heat engines have existed since the early eighteenth century, drove the Industrial Revolution and gave rise to the science of thermodynamics. Thermoelectricity was discovered about a century later[2] and is based on the same thermodynamic principles that heat engines depend on, except for the fact that thermoelectric power generators use electrons, rather than steam or air, as the working fluid. A testament to the importance of these fields is the fact that progress in the thermal sciences has been unrelenting: this includes work by Zhao *et al.*[3] on page 373 of this issue.

The second law of thermodynamics dictates that, to deliver work, heat engines must operate between a source of heat at a hot temperature (T_{hot}) and a heat sink at a cooler temperature (T_{cold}). The Carnot efficiency, $\eta_{max} = 1 - (T_{cold}/T_{hot})$, is the upper bound for the efficiency (η) of a heat engine, where η is the ratio between the amount of work an engine does and the amount of heat it uses. A thermoelectric generator works as follows. A temperature gradient, ∇T, across two thermoelectric materials — a semiconductor in which most of the charge carriers are electrons (n-type semiconductor) and a semiconductor that has mostly notional holes created by the absence of electrons (p-type) — creates an electric field, E, between the cold side and the hot side of each material (Fig. 1). The Seebeck coefficient, S, which is given by the ratio $E/\nabla T$ and is negative for the n-type material and positive for the p-type, corresponds[4] to the entropy of the electron divided by its charge. The two materials complete a cycle that converts the heat supplied at the hot side into electrical power. Assuming that S does not vary along the length of each thermoelectric material even though the temperature does, if this cycle were reversible it would have the Carnot efficiency.

Thermodynamically irreversible processes limit the efficiency of the cycle to a value much lower than that of the Carnot efficiency. Examples of such processes are heat conduction through the crystal lattice of atoms that constitute the semiconductors and the Joule heating that arises inside the semiconductors when the voltage produced by the electric field is used to deliver a current to an external electrical load (Fig. 1). The fraction of the Carnot efficiency of a thermoelectric cycle is quantified by the thermoelectric figure of merit ZT

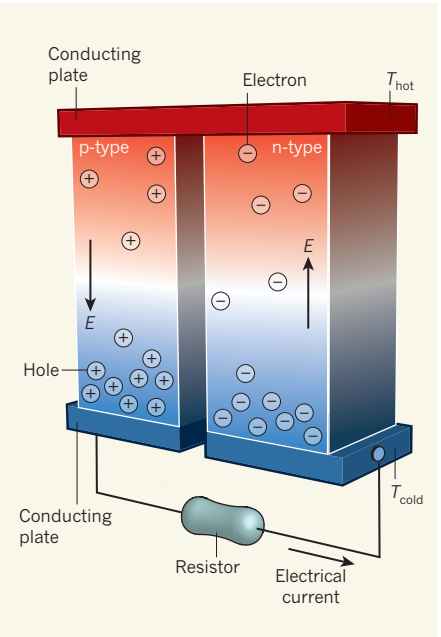


Figure 1 | Working principle of a thermoelectric generator. a, A thermoelectric generator consists of two thermoelectric semiconductors (n-type and p-type) subjected to a temperature difference, $T_{hot} - T_{cold}$, and electrically connected in series through conducting plates on the top and bottom. In the n-type semiconductor, most charge carriers are negatively charged electrons, whereas in the other one most of the carriers are positively charged holes. In a temperature gradient, electrons and holes tend to accumulate on the cold side. An electric field E develops between the cold side and the hot side of each material, which gives a voltage when integrated over the length of each. The voltages of the n- and p-type semiconductors add up and drive an electrical current through an electrical load, here an electrical resistor. The product of the voltage and the current is the electrical power output of the generator.

© 2014 Macmillan Publishers Limited. All rights reserved

of the system as a whole, which is the average of the zT of the n- and p-type semiconductors, where $zT = S^2T/\rho\kappa$, with ρ representing each semiconductor's electrical resistivity, κ its thermal conductivity and T the absolute temperature of the device. The thermal conductivity has two components: electrons carry some heat (electronic conductivity, κ_e), but most of the heat is carried by phonons, vibrations of the atoms in the crystal lattice that also transport sound (lattice conductivity, κ_{lat}). The goal of thermoelectric research is to discover new materials with maximum zT, by maximizing the ratio S^2/ρ and minimizing κ_{lat} without increasing ρ.

The maximum achievable value of zT has doubled in the past 15 years[5] from 1 to 2, thanks to the application of nanotechnology and quantum theory to this problem. Engineering the energy-band structure of the materials[5] — for example, through the effect of quantum confinement or by enhancing interactions between the wavefunctions of impurities and of free electrons — is used to boost S^2/ρ. Decreases in κ_{lat} can be achieved by phonon engineering[5] of the semiconductors — by, for instance, nanostructuring them, generating specific localized lattice vibrations (rattling phonon modes) in them, or selecting atoms that induce chemical bonds that vibrate highly anharmonically[6]. For decades before Zhao and colleagues' study, which examined the thermoelectric efficiency of single crystals of tin selenide (SnSe), the record for zT was held by alloys created using several of these techniques in lead telluride[7] (PbTe), a semiconductor that has a simple rock-salt crystal structure and a long history in thermoelectric technology, infrared diode lasers and radiation detectors.

In comparison to PbTe, SnSe is the ugly duckling. Chemically and structurally akin to PbTe, SnSe is lighter, has stiffer bonds and a distorted lattice. This made it seem a poor choice for thermoelectrics; indeed, SnSe has a low zT at room temperature. It was a surprise to learn that Zhao and colleagues had investigated SnSe at all, let alone to see it achieve the highest zT (2.6 along one crystallographic direction of its high-temperature phase) of any bulk material in an as-grown sample; that is, one without any addition of impurities or other optimization. SnSe has good prospects for practical use: it is not subject to the legislation that limits the use of Pb, contains only Earth-abundant elements (unlike Te), and can be prepared with good reproducibility[3].

Finally, the physics of SnSe is fascinating. The authors attribute the material's low κ_{lat} to the high anharmonicity of its chemical bonds. A solid with purely harmonic bonds would look like a three-dimensional array of balls and springs. If an atom is pulled from its equilibrium position during the passage of a phonon, the force that the atom is subjected to is proportional to its displacement, and the proportionality constant of this relationship is called the spring constant. In an anharmonic solid, the spring constant does not remain constant with atom displacements, which has important consequences when two phonons run into each other. The presence of the first phonon then changes the value of the spring constant seen by the second phonon. The second phonon thus runs into a medium with modified elastic properties, which is more likely to reflect it. Anharmonicity results in enhanced phonon–phonon scattering, which reduces κ_{lat} without affecting the solid's electronic properties[5,6]. Therefore, this effect may well be behind the high zT of SnSe, an idea that will stimulate further experimental and theoretical work. ■

Joseph P. Heremans *is in the Departments of Mechanical and Aerospace Engineering and of Physics, Ohio State University, Columbus, Ohio 43210, USA.*
e-mail: heremans.1@osu.edu

1. www.iea.org
2. Seebeck, T. J. *Magnetische Polarisation der Metalle und Erze durch Temperatur-Differenz* (Abhandlungen der Preussischen Akad. Wissenschaften) 265–373 (1822–23; reprinted W. Engelmann, 1895).
3. Zhao, L.-D. *et al. Nature* **508,** 373–377 (2014).
4. Callen, H. B. *Thermodynamics: An Introduction to the Physical Theories of Equilibrium Thermostatics and Irreversible Thermodynamics* (Wiley, 1960).
5. Heremans, J. P., Dresselhaus, M. S., Bell, L. E. & Morelli, D. T. *Nature Nanotechnol.* **8,** 471–473 (2013).
6. Nielsen, M. D., Ozolins, V. & Heremans, J. P. *Energy Environ. Sci.* **6,** 570–578 (2013).
7. Biswas, K. *et al. Nature* **489,** 414–418 (2012).

CELL BIOLOGY

The stressful influence of microbes

An investigation into cellular stress responses reveals how cell compartments called mitochondria use information about the surrounding metabolites and microorganisms to protect themselves from damage. SEE LETTER P.406

SUZANNE WOLFF & ANDREW DILLIN

Cellular organelles are simultaneously distinct from the rest of the cell and completely reliant on it for their identity and function. A poignant example of this nebulous individualism comes from the mitochondrion, a subcellular entity that began long ago as an independent organism but which, over millennia, has become increasingly dependent on the rest of the cell, and now serves as its energy-generating centre. Interactions between the mitochondrion and the cell provide this organelle with a direct connection to changes in the contents of its surroundings, allowing it to initiate defence mechanisms when things go awry. In this issue, Liu *et al.*[1] (page 406) report the effect of cellular and microbial metabolites on the initiation of protective reactions in mitochondria.

When changes occur that are stressful to mitochondria (for example, changes in their ability to import the molecules required for normal function), they activate a protective program called the mitochondrial unfolded protein response[2] (UPR^{mt}). This defence mechanism is designed to maintain normal mitochondrial function, and its activation is typically symptomatic of an imbalance in the cell. A diverse set of stimuli can activate the UPR^{mt}, including disruption of the proteins, import machineries and protease enzymes that help the mitochondrion to function[3]. Initiation of this defence begins with transportation of the transcription factor ATFS-1 — considered to be the major factor in mitochondrial protection against stress[4] — to the nucleus, where it upregulates expression of other genes involved in the response. By contrast, under non-stressed conditions, ATFS-1 is imported into mitochondria, where it is readily degraded by proteases[4].

Liu *et al.* report that when they induced the UPR^{mt} in the nematode worm *Caenorhabditis elegans*, by genetically or pharmacologically inactivating genes involved in normal mitochondrial function, the worms displayed an aversion to bacteria — their typical source of food. They hypothesized that UPR^{mt} induction might cause such a change in behaviour because the cell interprets the program's activation as symptomatic of a pathogenic attack or as a sign of poor nutrient availability (Fig. 1). If so, this could imply that variations in metabolite levels, which might be caused by the presence of pathogens or a lack of nutrients, play an integral signalling part in the activation of mitochondrial stress responses. These hypotheses are supported by previous observations suggesting a variety of methods by which bacteria are linked to and influence metabolism[5] and mitochondrial stress[6] in *C. elegans*.

Next, Liu and colleagues set out to define the intracellular metabolic pathways that are required for UPR^{mt} induction. Using

© 2014 Macmillan Publishers Limited. All rights reserved

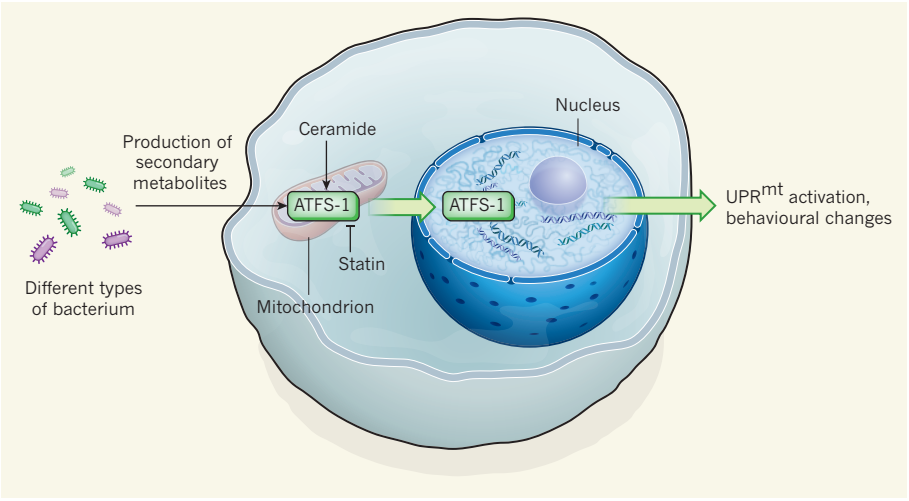


Figure 1 | Cellular and bacterial metabolites influence mitochondrial behaviour. When cellular mitochondria sense stressful changes in their environment, they activate a protective program known as the mitochondrial unfolded protein response (UPR^{mt}). This occurs owing to movement of the transcription factor ATFS-1 from the mitochondrion to the nucleus, where it activates UPR^{mt} genes, causing changes in cellular behaviour. Liu *et al.*[1] report that metabolic products secreted from bacterial cells can induce the UPR^{mt}. The authors find that the same response can be brought about by intracellular synthesis of metabolites such as ceramide or products of the mevalonate pathway, which is inhibited by statin drugs.

genome-wide screening, the authors identified 45 genes that are needed for the mitochondrial stress response. In this analysis, they found two metabolic pathways required for activation of the UPR^{mt}.

The first, the sphingolipid biosynthesis pathway, is responsible for production of ceramide, a waxy lipid molecule typically found in the cell membrane after being synthesized in another organelle, the endoplasmic reticulum (ER). Ceramide plays a part in multiple physiological processes, including mitochondrial degradation[7] and the cell-death program apoptosis[8]. Although the authors report that ceramide synthesis is required for UPR^{mt} induction, a loss of ceramide synthesis has also been linked to increased longevity in several organisms[9–11], suggesting a complex role for this molecule in health.

Disruption of a second metabolic pathway, the mevalonate pathway, also blocks activation of the UPR^{mt}. This pathway is required for the synthesis of cholesterol in most multicellular animals, but not in *C. elegans*[12], indicating that alternative metabolic outputs of this pathway (rather than cholesterol synthesis) are required for the mitochondrial stress response in this worm.

The mevalonate pathway is inhibited by statins, drugs that are used in humans to lower cholesterol. Previous work has demonstrated that statins have adverse effects on longevity in *C. elegans*, affecting the function of the ER and inducing a robust ER stress response[13], which acts to protect the ER in the same way as the UPR^{mt} protects mitochondria. Likewise, Liu and co-workers demonstrated that statins affect mitochondrial health. However, in contrast to what is seen in the ER, they found that after treatment with statins the worms could no longer respond to mitochondrial stress. In agreement with this observation, earlier data showed that forcing ATFS-1 to enter the nucleus protects worms from the harmful effects of statins[14], suggesting that a breakdown in mitochondrial stress sensing is responsible for the harmful effects of statins in *C. elegans*. This may have consequences for our understanding of the side effects of statins in humans.

As with statins, Liu *et al.* found that forcing ATFS-1 to move to the nucleus was sufficient to induce the UPR^{mt} when the sphingolipid biosynthesis pathway was blocked. Clearly, the typical regulation of ATFS-1 is broken after disruption of these two metabolic pathways. But what exactly prevents ATFS-1 from functioning properly?

The authors' data suggest that reducing mitochondrial import of ATFS-1 may not be sufficient to ensure its transportation into the nucleus. Alternatively, mitochondrial import and degradation of ATFS-1 may be directly controlled by metabolite changes in the cell. A closer look at the effect of these metabolites on ATFS-1 import and stability will offer a better understanding of how they affect the function of the UPR^{mt}. Moreover, these results suggest an effect of UPR^{mt} activation, and potentially of ATFS-1, on *C. elegans* behaviour in response to food sources. How this plays into the general sensing of food availability — for instance, through control of the nervous system[15] — is not yet known.

The extent to which this is a general phenomenon or one specific to these particular metabolites also remains unclear. The authors find that a remarkably high percentage of bacterial species in a large panel can induce the UPR^{mt} in *C. elegans* in the absence of other stressors. By contrast, they find that some classes of microbe are capable of blocking UPR^{mt} activation even during stress. The authors leave the mechanism behind these varying effects open, but the overarching message of this study is clear: different metabolites produced in these differing conditions are most likely to be responsible.

Mitochondria are ancient relics of a purely single-celled world. It is possible that the systems that protect mitochondria were the same as those used to protect colonies of bacteria from their invading neighbours. Perhaps bacterial metabolites have retained the ability to communicate with their long-lost relative, the mitochondrion, which in turn has gained the ability to communicate with its host to alter such complex processes as behaviour. Human physiology also relies on complex interactions with thousands of species of bacterium. It is possible that our own mitochondria will sense and respond to the secondary metabolites produced by these species, and possibly change the behaviour of our cells. ■

Suzanne Wolff *and* **Andrew Dillin** *are in the Department of Molecular and Cell Biology and the Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California 94720-3200, USA.*
e-mail: dillin@berkeley.edu

1. Liu, Y., Buck, S. S., Breen, P. C. & Ruvkun, G. *Nature* **508,** 406–410 (2014).
2. Zhao, Q. *et al. EMBO J.* **21,** 4411–4419 (2002).
3. Haynes, C. M., Petrova, K., Benedetti, C., Yang, Y. & Ron, D. *Dev. Cell* **13,** 467–480 (2007).
4. Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. *Science* **337,** 587–590 (2012).
5. Virk, B. *et al. BMC Biol.* **10,** 67 (2012).
6. Pang, S. & Curran, S. P. *Cell Metab.* **19,** 221–231 (2014).
7. Sentelle, R. D. *et al. Nature Chem. Biol.* **8,** 831–838 (2012).
8. Stiban, J., Caputo, L. & Colombini, M. *J. Lipid Res.* **49,** 625–634 (2008).
9. D'Mello, N. P. *et al. J. Biol. Chem.* **269,** 15451–15459 (1994).
10. Huang, X., Liu, J. & Dickson, R. C. *PLoS Genet.* **8,** e1002493 (2012).
11. Mosbech, M. B. *et al. PLoS ONE* **8,** e70087 (2013).
12. Kurzchalia, T. V. & Ward, S. *Nature Cell Biol.* **5,** 684–688 (2003).
13. Morck, C. *et al. Proc. Natl Acad. Sci. USA* **106,** 18285–18290 (2009).
14. Rauthan, M., Ranji, P., Pradenas, N. A., Pitot, C. & Pilon, M. *Proc. Natl Acad. Sci. USA* **110,** 5981–5986 (2013).
15. Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. *Nature* **404,** 661–671 (2000).

This article was published online on 2 April 2014.

CORRECTION

The News & Views article 'Tuberculosis: Drug discovery goes *au naturel*' by Clifton E. Barry (*Nature* **506,** 436–437; 2014) incorrectly described the pharmaceutical company Lepetit as being French. It is an Italian firm.

© 2014 Macmillan Publishers Limited. All rights reserved

REVIEW

doi:10.1038/nature13001

The ensemble nature of allostery

Hesam N. Motlagh[1], James O. Wrabl[1], Jing Li[1] & Vincent J. Hilser[1]

Allostery is the process by which biological macromolecules (mostly proteins) transmit the effect of binding at one site to another, often distal, functional site, allowing for regulation of activity. Recent experimental observations demonstrating that allostery can be facilitated by dynamic and intrinsically disordered proteins have resulted in a new paradigm for understanding allosteric mechanisms, which focuses on the conformational ensemble and the statistical nature of the interactions responsible for the transmission of information. Analysis of allosteric ensembles reveals a rich spectrum of regulatory strategies, as well as a framework to unify the description of allosteric mechanisms from different systems.

Allostery, first articulated more than 50 years ago[1–3], has remained a central focus in biology, owing to the fact that a quantitative description of allostery is fundamental to an understanding of most processes beyond the molecular level, such as cellular signalling[4] and disease[5]. Allostery has even been referred to as the 'second secret of life'[6,7]. In spite of its importance, allosteric mechanisms in most instances remain a biophysical enigma, eluding a general, quantifiable and predictive atomic description[7,8].

Through the lens of history, it is clear that allosteric concepts and models have evolved over decades as the experimental technologies have improved. The first allosteric proteins studied, such as haemoglobin[9], were extensively structured and thus amenable to high-resolution X-ray crystallography. Models of allosteric mechanisms based on such studies were understandably greatly influenced by the static structural images[10,11]. However, as more experimental structural, dynamic and thermodynamic approaches have been developed, descriptions for allostery have necessarily become more complex. The largely qualitative, static images of end point protein structures have been replaced by more quantitative, dynamic views of allostery, which, unlike static structural models, are more difficult to visualize. Here the historically dominant paradigm of structure-based allostery[11–14] is briefly reviewed and placed in the context of new discoveries, revealing that allostery can also be associated with changes in dynamics and large-scale conformational disorder[15–18]. Indeed, more recent discoveries emphasize how nature uses the full spectrum of conformational heterogeneity[19] (for example, rigid body movement[20–22], folded yet dynamic structure[4,23–27], and intrinsic disorder[28–30]) to facilitate allostery. These recent discoveries challenge traditional understanding of allostery and provide the foundations for developing a framework that unifies descriptions of allostery in structured, dynamic and disordered systems.

From structures to ensembles

The first statement of allostery[31], arising from the classic experiments of Changeux[2], is that two distinct sites within one protein, each binding different ligands, could nonetheless interact despite being 'non-overlapping' (that is, presumably distant from each other) in the molecular structure. This 'action at a distance' phenomenon was difficult to interpret in the absence of structural information or site-specific biochemical data.

For decades, the two dominant models for allostery were the 'sequential', or KNF (Koshland–Nemethy–Filmer), model[32] and the 'symmetric', or MWC (Monod–Wyman–Changeux), model[3]. Both proposals were influential and agreed on the importance of conformational change between two well-defined structural end states[33]. Whereas the MWC model posited the existence of two pre-existing quaternary states (tensed (T) and relaxed (R)), whose equilibrium was shifted upon binding ligand[3], the KNF model was based on the appealing, general notion of the inherent flexibility of proteins[32] and 'induced-fit' of a binding site in response to ligand[34,35]. Importantly, both models were successful in describing allostery[31–36].

However, it is well known that both the MWC and KNF models are phenomenological[33], and consequently, do not provide insight into how the structure facilitates allosteric communication between sites. It was the birth of structural biology and the development of the influential stereochemical model by Perutz[10,13] that first addressed this issue and set the course for future allosteric studies. According to this model, which was built upon by many others[33,37,38], allostery could be understood in terms of structural changes that could be gleaned through inspection of the high-resolution structure. This structural view of allosteric mechanism has been the reigning paradigm ever since[39–43], with some approaches even positing the existence of conserved allosteric pathways that presumably mediate the transmission of information between sites. In fact, such pathways are even evident in haemoglobin[44] (Fig. 1a).

Nonetheless, the notion that something was amiss (or at least incomplete) with a purely structure-centric understanding of allostery had been recognized for decades. In a classic thermodynamic study of haemoglobin, Parsegian and colleagues[45] found that the T-to-R transition in haemoglobin involved the binding of approximately 60 additional water molecules (Fig. 1b), and noted "… it is doubtful that the R state would even exist without the stabilization promoted by hydration of the extra surface." In effect, the entire molecule must have a role in mediating allostery. Contrasting the different 'images' of allostery conveyed by Fig. 1 reveals the challenge in quantitatively describing mechanism. Clearly, an understanding of allostery requires not only knowledge of the concerted structural changes between two binding sites, but also the factors that render an allosteric protein 'poised to undergo' the structural change.

The Parsegian quote itself hints at the realization (well-appreciated by the authors of the early models) that allosteric transitions involve multiple states, whose relative populations must be tunable over a range of physiological conditions. Nonetheless, the fact that initial structures were found for the T and R forms of haemoglobin supported the notion that allostery as a general phenomenon might be understood, perhaps in all systems, if only the structures of the allosteric states could be determined. To that end, extensive molecular dynamics[33,46,47] and coarse-grained network simulation[20,48] (for those proteins in which rigid body motions could be proposed for the allosteric transition) further suggested that 'allosteric mechanisms reliant primarily upon observable changes in the ensemble-average structure are common and may be dominant'[20]. Although the discovery of an alternate 'relaxed' structure of haemoglobin by Arnone and co-workers[49] laid to rest the simple two-state view of allostery, at least for haemoglobin, not fully appreciated was the range of possible regulatory

[1]Department of Biology and T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA.

©2014 Macmillan Publishers Limited. All rights reserved

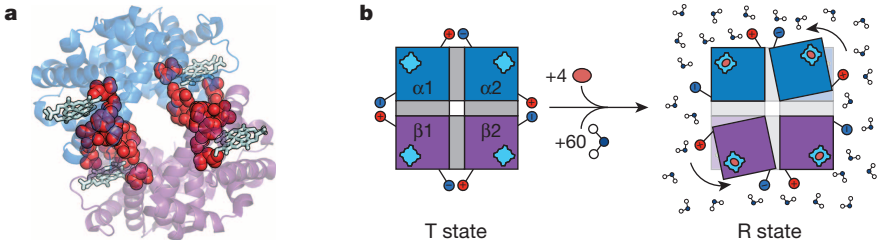


Figure 1 | Structure-based views of allostery. **a**, Ribbon diagram representation of tetrameric haemoglobin (PDB accession 1GZX) rendered in PyMol (Schrödinger). The proposed pathway responsible for the cooperative transition from tensed (T) to relaxed (R) is highlighted with red spheres and the haem groups are represented as light blue sticks[44]. **b**, Allosteric transition of tetrameric haemoglobin, as proposed by Perutz[10,13]. Tetrameric haemoglobin in the T state is depicted on the left with the two α-subunits (blue) and the two β-subunits (purple) each with their own haem group (light blue). Salt bridges, depicted as the red positive and blue negative charges, hold the molecule in the T conformation, and these salt bridges are released upon binding of oxygen (orange oval) in the transition to the R conformation (on the right) accompanied by a 15° turn of the subunits relative to each another. Also contributing to the equilibrium are 60 additional water molecules preferentially binding the R state[45].

strategies that could be used if the full spectrum of binding and conformational states of proteins were to be harnessed.

In 1984 Cooper and Dryden touched on one such possibility when proposing that allostery could be manifested from changes in the width of a conformational distribution, raising the spectre of significant entropic contributions to allostery[50]. These authors rigorously demonstrated, using a statistical thermodynamic formalism, that changes in the frequency and amplitude of thermal fluctuations in a protein upon ligand binding could result in cooperative energies on the order of a few kcal per mol without perturbing the average structure. Importantly, such a situation would be difficult to reconcile by simple inspection of the end-state structures, explaining why experimental examples had escaped detection until recently. The work of Cooper and Dryden has proven prescient[33]. The term 'dynamic allostery' has even found its way into the modern lexicon, ostensibly to describe the role of entropy in the thermodynamics of allostery, although unfortunately it is often conflated with motions between relevant ensembles. However, with the development of NMR techniques to study allosteric protein systems at a site-resolved level, both aspects of the role of motion in transitions between functional states are becoming illuminated[51–55].

One of the most surprising observations over the past half-decade has been the appearance of allostery in intrinsically disordered proteins (IDPs), systems that are defined by a lack of fixed structure, exhibiting instead extreme conformational fluctuations[56–60]. Clearly both conformational dynamics of folded structures and large-scale disorder are important for allostery[61], but quantitatively understanding this phenomenon remains a challenge. It is not obvious how the phenomenon of allostery can be understood and described in terms that can do equal justice to both highly structured and highly disordered systems.

Below, we use several recently studied allosteric systems, spanning the continuum of structure/dynamics classification space (Fig. 2), to show how changes in dynamics and disorder are phenomenologically associated with allostery. Initially highlighted are the challenges associated with reconciling allostery in terms of classic structural or deterministic models. However, when allosteric proteins are viewed as ensembles of states, an entirely new palette of expected behaviour is revealed. Importantly, it becomes clear that taxonomic schemes segregating allosteric proteins by amount and/or type of dynamics that change during an allosteric transition may actually obscure similarities between mechanisms that are only phenomenologically different (Box 1). The ensemble nature of allostery revealed through analysis of the full dynamic spectrum of allosteric systems permits a rich repertoire of regulatory strategies and illuminates key organizing principles for describing allostery.

The dynamic continuum of allostery

Arguably the most significant experimental advance in the analysis of allostery has been the development of NMR spectroscopy, which provides information about protein structure, dynamics and thermodynamics[16,53,54]. Although the high-resolution structure of a protein provides exquisite insight into the enthalpic contributions of structure and solvent to protein thermodynamics[62], the role of entropy remains obscure. The idea of using fast internal motion detected by solution NMR as a proxy for conformational entropy was introduced some time ago[63,64], where motion evaluated assuming a specific motional model provided an indirect measure of the number of states visited. The detection of motion in proteins over a wide range of timescales is now well established, but it is the fast picosecond-to-nanosecond timescale that seems to be most important for the detection of conformational entropy[65–67]. The initial model-dependent interpretation of protein dynamics in terms of entropy was problematic[65] but its use did reveal the probable presence of a large contribution from conformational entropy to the free energy of protein–ligand interactions[68,69]. To overcome many of the technical difficulties associated with using a dynamic proxy for conformational entropy, Wand and co-workers developed an empirical 'entropy meter' that appears able to quantitatively relate changes in fast internal protein motion measured by NMR relaxation to changes in conformational entropy[54,70]. This approach has been used to demonstrate the integral role of conformational entropy in molecular recognition by calmodulin[70] and the catabolite activator protein (CAP)[24]. The calmodulin

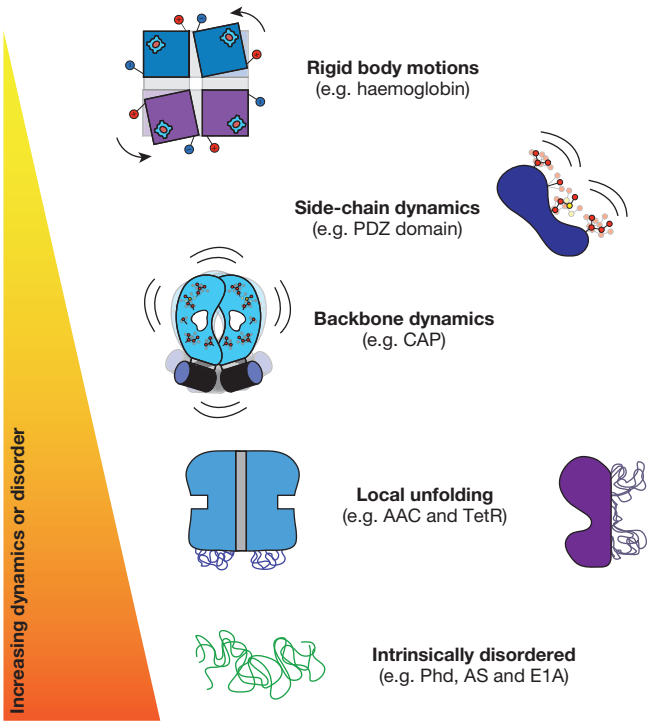


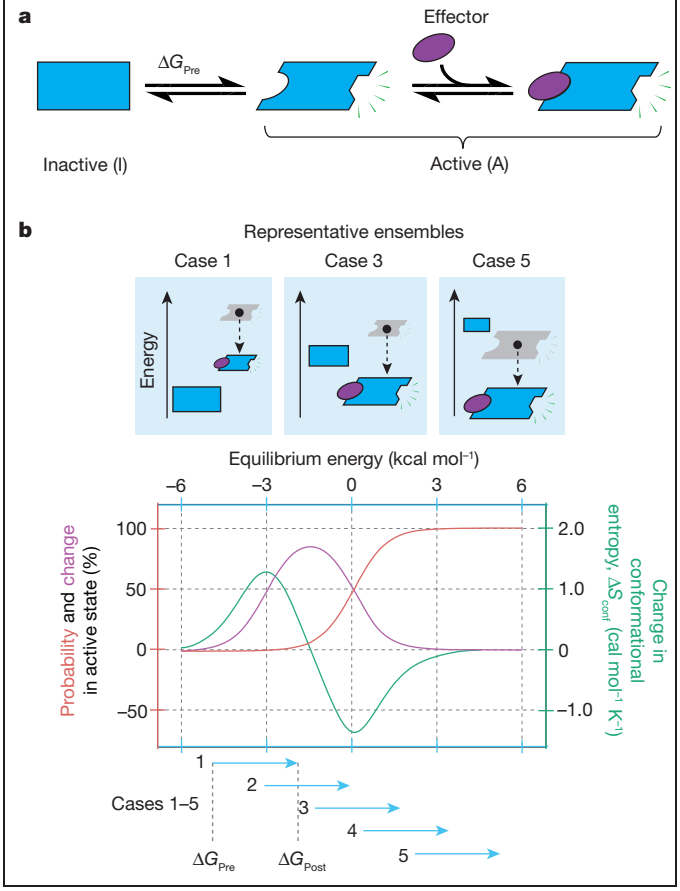
Figure 2 | The dynamic continuum of allosteric phenomena. Schematic representation of allosteric systems with increasing dynamics, disorder or fluctuations on the vertical axis.

©2014 Macmillan Publishers Limited. All rights reserved

BOX 1

Tunable sensitivity in allosteric ensembles

Tuning the sensitivity of the ensemble to an allosteric effector. **a**, Two-state allosteric protein wherein the effector ligand binds to and stabilizes the active state. This equilibrium (ΔG_{Pre}) can be shifted (ΔG_{Post}) by adding effector. **b**, Shown is the response of the system to an allosteric effector that increases the stability of the active state by 3 kcal mol^{-1}. The observed response is dependent on where the equilibrium is poised before activation. When the inactive state is too dominant (**b**, top, case 1), activation (that is, adding effector) produces a modest change in activity and an increase in the conformational entropy, S_{conf} ($= -R\Sigma P_i \ln P_i$, where P_i is the probability of the ith state and the summation is over all states in the ensemble). When the active state is already dominant (case 5), activation again produces a modest effect, but in this case the conformational entropy decreases. Under conditions in which both states are populated (case 3), the change in activity is maximal, but the change in conformational entropy is near zero. This is depicted quantitatively in the graph in the bottom of **b**: the probability of being active (red), the change in activity (purple) and the change in conformational entropy (green) as a function of ΔG_{Pre}. Importantly, the free energy before (ΔG_{Pre}) and after (ΔG_{Post}) addition of effector ligand determines how the ensemble is poised, which in turn determines both the degree of activity change and whether conformational entropy (as determined from a positions-specific probe of dynamics) will increase, decrease or remain constant upon activation.



system provided illustrations of the various types of dynamical perturbation that can occur: purely local, long-range or directed dynamical transmission, which is perhaps the key to dynamically mediated allostery[71].

The CAP system has been particularly important in exploring the role of conformational entropy in allostery, as discussed below.

NMR analyses of allosteric proteins have revealed a cornucopia of ways that changes in dynamics are associated with allosteric transitions. Indeed, a continuum of dynamic allosteric phenomena has been observed, ranging from rigid body structural changes to intrinsically disordered proteins mediating the response (Fig. 2). The existence of this dynamic continuum suggests that although the two extremes of the spectrum represent radically different structural 'pictures' of allostery, they should nonetheless be amenable to a unified description.

Structured yet dynamic allosteric systems

The classic view of an allosteric protein is one of structure. The protein may have one or more domains or subunits whose structure can be determined, but which may show conformational dynamics when observed in NMR experiments. Many examples of structured allosteric proteins that show dynamic changes have been documented[4,18,23–26,72]. For reasons described below, these are ideal systems to probe ensemble properties. For now we simply note that the analyses of these types of systems have revealed a number of interesting paradoxes.

Since the initial prediction of dynamic allostery without conformational change by Cooper and Dryden[50], two important experimental studies have gone on to show that allostery can indeed be manifested essentially without structural change. The first study focused on CAP, a homodimeric transcription factor consisting of a cyclic AMP (cAMP) binding domain coupled to a DNA binding domain (DBD) (Fig. 3a). The biological role of CAP is to bind cAMP, allowing CAP to activate transcription (Fig. 3a, right). Extensive studies have revealed that conformational entropy via backbone and side-chain dynamics is associated with the allosteric response of the system upon cAMP binding[24–26]. Whether the entire entropic penalty arises from quenching of dynamics upon ligand binding is an open question[33], but certainly a significant contribution of the allosteric response is associated with the quenching of dynamics. Interestingly, high-resolution NMR spectroscopy has revealed that the population of binding-competent states for wild-type and mutants of CAP in solution is not proportional to the overall binding affinity. Instead, the entropic effects for a number of mutants were shown to correlate with the conformational entropy determined from NMR order-parameters[24]. In this case, the entropic penalty seems to be the basis of the negative cooperativity—binding cAMP increases dynamics in the adjacent domain that must be quenched to bind the second ligand. Interestingly, a recent study of CAP using elastic network models demonstrated that modulation of the normal modes can reproduce these results and can be used to rationally modulate the cooperativity of ligand binding[73].

In a second important study, removal of an α-helix (Δα3) from a canonical PDZ domain resulted in an increase in side-chain dynamics of the distal binding site[23] (Fig. 3b). The helix in question was presumed to be a molecular scaffold and not expected to affect binding, because it was not present in most other PDZ domains and was physically distal from the binding site[23]. Unexpectedly, deletion of α3 (depicted blue in Fig. 3b, right) resulted in a 25-fold reduction in binding affinity that was almost entirely entropic in nature, as determined by isothermal titration calorimetry. NMR studies revealed no global structural perturbations, and relaxation experiments showed no significant changes in the backbone dynamics. The only detectable differences were observed in the side-chain dynamics (Fig. 3b, top right) and the calorimetrically determined entropy change (in other words, the dynamics in the unliganded state). Both the CAP and the PDZ examples were pioneering studies that not only revealed the predicted importance of dynamics and the role of conformational entropy in allosteric mechanism, but also highlighted the limitations of deducing mechanism from static structure alone.

Local unfolding and intrinsic disorder

Moving down the dynamic continuum (bottom of Fig. 2), one of the more recent developments is the realization that local unfolding and intrinsic disorder[61] can have a significant role in allosterically coupling domains to

©2014 Macmillan Publishers Limited. All rights reserved

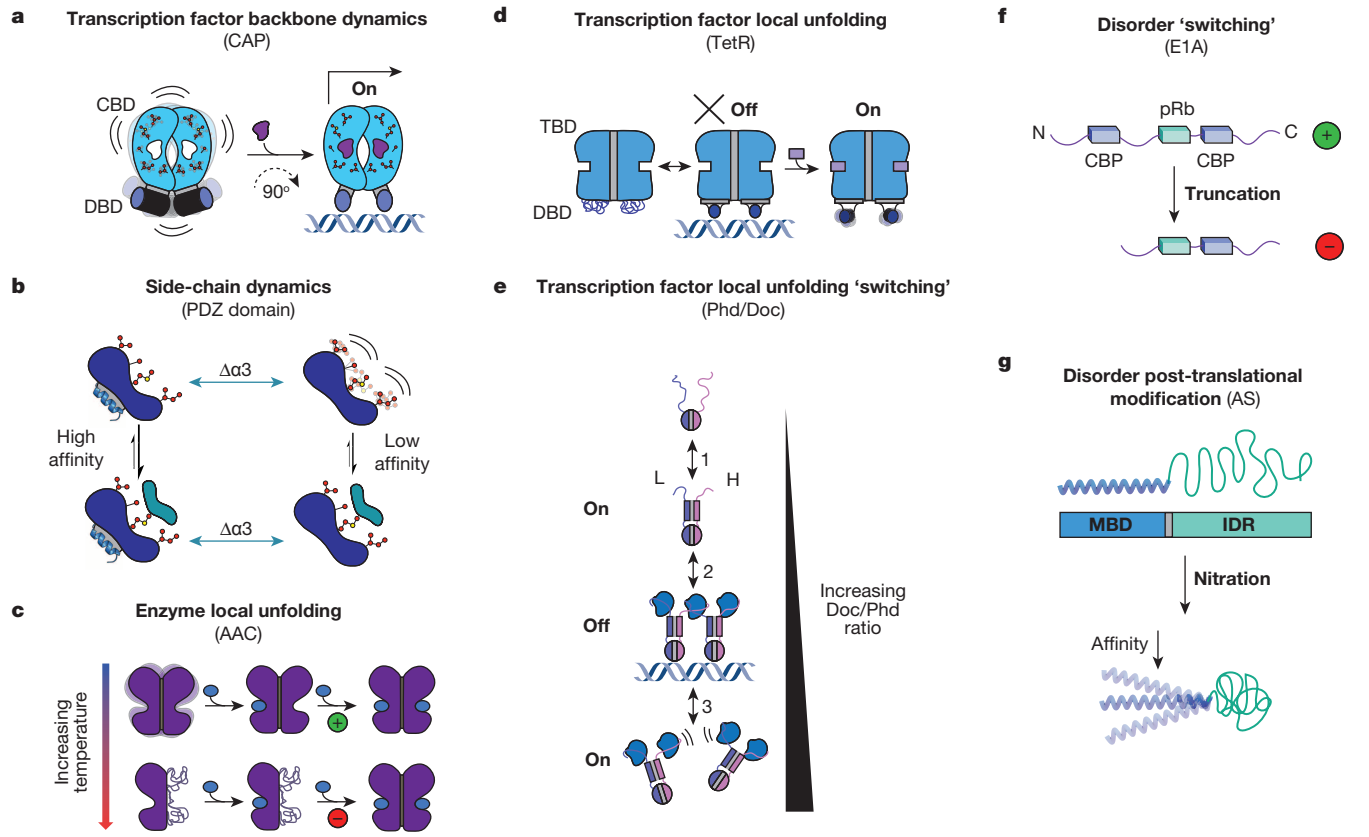


Figure 3 | Allosteric systems from the dynamic continuum. **a**, CAP homodimer with a cyclic nucleotide binding domain (CBD, blue domains with side chains) and a DNA binding domain (DBD, blue cylinders). Binding energetics of cAMP (purple ligand) quench dynamics in the bound-state ensemble and induce a 90° change in the conformation of the DBD, allowing it to bind DNA and turn on transcription[24–26]. **b**, Side-chain dynamics modulate the binding affinity of a canonical PDZ domain to its native ligand with and without α-helix 3 (Δα3, green arrows)[23]. **c**, AAC homodimer with each monomer represented as one purple domain. Binding of the allosteric effector acetyl-CoA (blue oval) is positively cooperative at low temperatures (green '+'), and negatively cooperative at high temperatures (red '−')[4]. **d**, TetR homodimer[27] depicted as a two-domain protein with a tetracycline binding domain (TBD, blue region, top) and a DBD (blue region, bottom). **e**, Doc/Phd toxin–antitoxin system equilibrium. Phd is depicted as a homodimer (top, blue and purple monomers) and Doc is depicted as the blue ligand[28]. **f**, Representation of variants of E1A[30] shown from N to C terminus with binding sites for ligands in the two variants represented by blue rectangles (CBP) and green rectangles (pRb). **g**, Schematic representation of α-synuclein (AS) with its N-terminal membrane-binding domain (MBD) coupled to its C-terminal IDR. Upon oxidative stress (nitration) the affinity of the MBD decreases (bottom)[29].

one another. As shown below, the hallmark feature of these systems is the apparently complex 'tunable' coupling between sites. An example in which local unfolding has a role in allostery is in the enzyme aminoglycoside *N*-(6′)-acetyltransferase II (AAC)—a homodimeric enzyme from *Enterococcus faecium* responsible for conferring bacterial resistance to aminoglycoside antibiotics[4]. An allosteric effector of AAC, acetyl-CoA, binds with positive cooperativity at low temperatures and switches to negative cooperativity at higher temperatures as determined by isothermal titration calorimetry (Fig. 3c). NMR reveals that certain backbone amides undergo dynamic broadening as a function of temperature, suggesting that local unfolding is responsible for this switching behaviour. This observation agrees with the change in enthalpy of binding as temperature is increased; the nonlinear temperature dependence at higher temperatures being a signature of local unfolding[74].

Another example in which local unfolding has a role in allosteric coupling is the tetracycline repressor (TetR), a homodimeric transcription factor consisting of a Tet binding domain (TBD) and a DBD (Fig. 3d). TetR binds DNA in the absence of the antibiotic (Tet) and is released when Tet binds the TBD, inducing a reduction in binding affinity of the DBD for DNA. A paradoxical observation from structural data was that the inactive Tet-bound (holo) form shows the DBD α-helices, which are involved in binding DNA, to be in a more binding-competent conformation than when they are in the active apo-form. Chaotropic denaturation and *in vitro* operator binding assays performed on TetR, as well as biologically relevant TetR mutants, revealed that local unfolding of the DBD in the native state ensemble was responsible for this paradox[27]. The protein functions because the DBD is locally unfolded in the absence of Tet, allowing enough conformational flexibility to bind DNA, and hence an increased binding affinity in the absence of Tet (Fig. 3d, left equilibrium). Upon binding Tet, the α-helices responsible for DNA binding are redistributed into an ensemble of conformations that are not completely binding competent. Interestingly, naturally occurring mutants exist that are Tet insensitive and/or have inverted responses, manifesting complexity and tunability like that seen in allostery of locally unfolded or intrinsically disordered sequences[27].

Perhaps one of the fastest evolving fields in biophysics is protein intrinsic disorder[59,60]. IDPs stand as a challenge to the structure–function paradigm, as IDPs have no stable tertiary structure under physiological conditions yet are able to function[58–60]. Seminal bioinformatics and experimental work has revealed that intrinsically disordered regions (IDRs) and IDPs are hyper-abundant in transcription factors and cell signalling pathways. Aside from the fact that they lack a stable structure, IDPs nonetheless possess features that would otherwise make them suitable allosteric proteins. These include (1) high specificity for multiple targets resulting from their structural plasticity[30]; (2) high specificity and low-affinity binding[18]; (3) fast association and dissociation rates for proper turnover[58]; and (4) enrichment of post-translational modification sites[75] and splicing isoforms[76]. To no surprise, several IDPs have been shown to exhibit allostery as part of

©2014 Macmillan Publishers Limited. All rights reserved

their role in signalling. Interestingly, just as with proteins that exhibit local unfolding, allosteric IDPs exhibit complex coupling behaviour. Three examples highlight this complexity.

The first case involves the Phd/Doc toxin–antitoxin system (Fig. 3e) that inhibits the ribosome A site. As part of its regulatory cycle, Phd controls its own transcription by binding DNA at its own operon. Notably, increasing the ratio of Doc/Phd initially inhibits transcription, and as the ratio exceeds a certain biological threshold, transcription is activated again, exhibiting 'conditional cooperativity'[28]. Conditional cooperativity results from the delicate balance of the ensemble of states present in solution. Phd has a locally unfolded carboxy terminus, which undergoes a disorder-to-order transition that exposes high-affinity (H) and low-affinity (L) binding sites (equilibrium 1, Fig. 3e). Increasing the ratio of Doc/Phd leads to partial saturation of the H and L sites by Doc, which in turn stabilizes the DBD of Phd, allowing DNA binding and transcription repression (equilibrium 2, Fig. 3e). Finally, increasing the Doc/Phd ratio further leads to saturation of the H and L binding sites, which leads to dissociation from DNA and activates transcription again (equilibrium 3, Fig. 3e).

Another recently reported example[30] involves the IDP adenovirus early region 1A protein (E1A) (Fig. 3f). E1A can bind two different host proteins, CREB binding protein (CBP) and retinoblastoma protein (pRB), in its capacity to disrupt host cellular signalling. Single-molecule fluorescence resonance energy transfer (smFRET) was used to dissect the cooperative nature of the coupled folding and binding involving CBP and pRB. By measuring the cooperativity of binding using different truncated E1A constructs, these authors elegantly demonstrated cooperativity switching between the CBP and pRB sites, wherein positive coupling (green '+' in Fig. 3f) becomes negative (red '−' in Fig. 3f).

Importantly, IDPs are enriched in post-translational modifications[75], telling of separate mechanisms that nature may use to regulate the allosteric response of the ensemble. An example of this occurs in α-synuclein, implicated in Parkinson's disease[29]. α-Synuclein is composed of an amino-terminal membrane-binding domain and a C-terminal IDR (Fig. 3g) and exhibits reduced membrane-binding affinity under oxidative stress[29]. Using smFRET and NMR, it was shown that nitration of the C terminus shifts its conformational ensemble to a more globally extended conformation, which is accompanied by a reduction in binding affinity (Fig. 3g, bottom). Because the C-terminal IDR is known to interact with a variety of ligands, these results suggest that α-synuclein may use this allosteric coupling to regulate the biological response. Modulation of the conformational ensemble in IDPs and IDRs via these post-translational modifications may be a general mechanism for rapid, reversible control of biological response. How this may occur is discussed in more detail in the following section.

As the examples in Fig. 3 reveal, the hallmark feature of allostery in locally unfolded and intrinsically disordered proteins is a general mechanism whereby cooperativity between two sites can not only be tuned (high or low) by environmental conditions, but the actual sign of the coupling can change, transforming an activator into a repressor, or vice versa. A single static picture of the protein does not reveal how the same ligand, binding to the same site on the same protein, can transmit structural information from one site to another, causing a more favourable conformation to bind ligand under one circumstance and a less-favourable conformation under another. Restated more bluntly, it is not clear how the same ligand binding to the same site can cause either one biological effect or its 'opposite'.

Here we show that by formulating the phenomenon of allostery in terms of thermodynamic ensembles, many of the seemingly complex and confounding properties of allosteric proteins, including the results of adding different allosteric ligands, emerge as a straightforward consequence of the heterogeneity in the system. From these allosteric ensembles, ground rules emerge that can be used to guide future experimental inquiry.

Allosteric ligands remodel the energy landscape

Much in same way that the ability of a protein to fold can be understood in terms of the conformational 'energy landscape' of the protein[77–79], so too can allosteric coupling. According to this framework, all possible conformations of a protein are sampled (or populated) according to their energies; lower energy conformations are sampled more often than those of higher energy. Because the binding of any ligand to a state increases how often the state is sampled[15,80], allosteric ligands effectively remodel the energy landscape of allosteric proteins[18]. This key point has been discussed, at least in qualitative terms, in a number of recent reviews[5,15,18]. The take-home message is that the energy landscape can be smooth with many accessible states in its conformational ensemble, discrete with few states, or something in between, and each of these scenarios can provide its own biological advantage.

However, stabilities and coupling energies in proteins are generally very modest, suggesting that no single state, or sets of allosteric states, will necessarily dominate the ensemble. In other words, allosteric mechanisms may be more statistical, and less deterministic, than classical models suggest. Indeed, a more quantitative look at the landscape, using ensemble models that divide the protein into interacting component parts[61,81,82], illuminates the statistical nature of the signalling process, and reveals a wealth of regulatory opportunities directly resulting from the heterogeneity in the ensemble (Box 2).

Of course, treating allosteric proteins as ensembles is not a new idea. Early work on the lambda phage repressor, for example, demonstrated that consideration of all the accessible ligated states was necessary to understand its biological function[83]. However, when conformational heterogeneity at each binding site is explicitly considered in addition to the degeneracy of ligated species, important ground rules for allostery emerge[8].

First, all ensembles have tunable sensitivity, wherein the same perturbation can produce a range of results, simply based on where a given equilibrium is poised before perturbation. This reality not only affects function but also the interpretation of conformational entropy changes in allosteric transitions, which, in principle, can change from positive to negative, even for the same transition (Box 1).

Second, ensembles contain a plurality of potential allosteric mechanisms, a point that can be demonstrated by noting that allosteric proteins often segregate different functions (for example, effector-binding and activity) into different domains. This allows for a straightforward enumeration of the ensemble that accounts for each functional domain to be in the active or inactive conformation[8,61,84], both with and without other domains in their functional states. Binding of the effector molecule stabilizes the active form of the effector-binding domain, which in turn can stabilize (or destabilize) the functional domain (Box 2)[61]. This is so because, according to landscape theory[77–79], all of the possible states in the ensemble have some finite probability based on their energies. Thus, if landscape theory holds, within any ensemble, there exist sub-ensembles in which two different binding sites are either positively coupled (both domains active) or negatively coupled (one domain active and the other inactive) (Box 2). In other words, the potential for different allosteric mechanisms is already pre-encoded in the ensemble, a property that is uniquely provided by an ensemble allosteric model[8].

The question remains as to what determines positive or negative coupling between two binding sites. In the context of the ensemble, the answer is the relative stability of the different states, which is determined by the equilibrium between the active and inactive state of each domain and the interaction energy between domains. In fact, the signs of the interaction energies are the key parameters in determining whether allosteric proteins can switch between activation and repression[61,84]. The critical insight is that the overall allosteric behaviour is an ensemble-weighted contribution of all the states and thus cannot be reconciled in the context of a single structure[85]. Instead the effect is statistical. Under one set of conditions, the ensemble could be poised such that two domains would be positively coupled, because the properties of the positively coupled sub-ensemble dominate the average. However, any perturbation of the ensemble that remodels the energy landscape may change which sub-ensemble dominates, and thus, transform activation into repression, or vice versa (Box 2).

Importantly, perturbations can arise from any number of sources, including post-translational modification (for example, phosphorylation, methylation, sumoylation, nitration[29]), alternative splicing or translation start sites[76], truncation[30], temperature[4] and the binding of a second allosteric

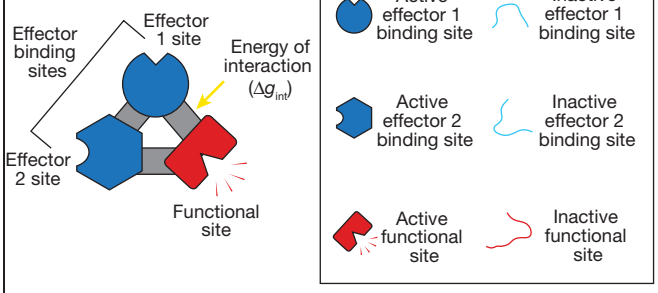
©2014 Macmillan Publishers Limited. All rights reserved

BOX 2

The ensemble view as a framework to explain allostery

a–**c**, Allosteric proteins can be viewed, to a first approximation, as a functional domain coupled to one or a number of effector-binding domains that interact with each other. To illustrate how an ensemble can switch the outcome of binding a particular effector molecule, we consider a protein with two effector-binding domains, which can interact with each other, as well as the functional site (**a**). Each domain can be inactive or active, together with (or separately from) the other domain, producing an ensemble of eight possible states (**b**, first column). In each state, the effector 1 binding site is in the active conformation (**b**, third column) either with or without the functional site in the active conformation (**b**, second column). Thus, the ensemble consists of states in which effector 1 site is positively coupled to the functional site (states 1 and 3; **b**, fifth column) and negatively coupled (states 4 and 5; **b**, sixth column)—that is, the effector 1 and functional sites are both active, or the effector 1 site is active while the functional site is inactive, respectively. Which set of states will govern the overall observed allosteric transition is determined primarily by the magnitude and sign of the interaction energies, Δg_{int}, between each domain (**a**)[61]. To see how this affects function, we consider the energy of all the states in the ensemble under a particular set of conditions (**c**)[8]. In the absence of effector 1, protein conformations in which the functional site is active have high energy, and so the probability of the functional site being active is low. Effector 1 binding lowers the energies (dotted lines in **c**) of all states in which the effector 1 binding site is active, including conformations that contain the active form of the functional site. This increases the probability that the functional site will be active. The sizes of the depicted structures indicate the probability of their formation; larger structures are more likely to form. Addition of effector 2, in the absence of effector 1, redistributes the ensemble to a new starting position (**d**, left) so that the two lowest energy conformations now have active functional sites. Addition of effector 1 binds to and lowers the energy of the same states as described above (dotted lines in **d**). In this instance, however, only one of the two lowest energy states has the functional site in the active conformation, lowering the probability that the functional site will be active. As a result, effector 1 switches from being an activator to a repressor[84].

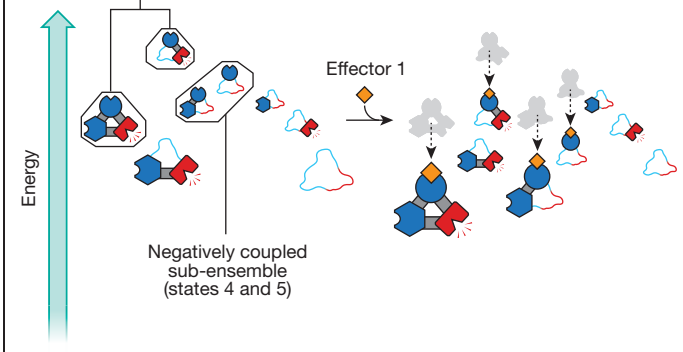
a



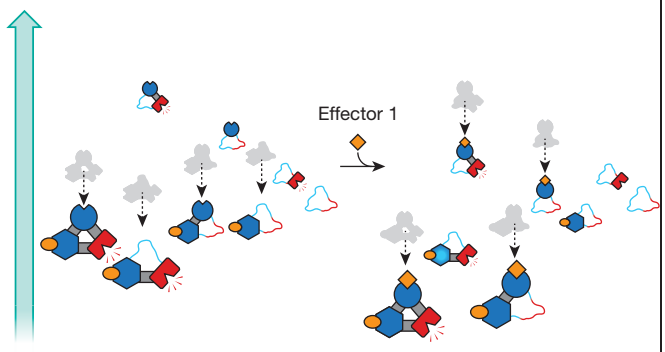
b

State	Active functional site	Active effector 1 site	Active effector 2 site	Coupling of effector 1 site and functional site: Positive	Coupling of effector 1 site and functional site: Negative
1	✓	✓	✓	✓	
2	✓		✓		
3	✓	✓		✓	
4		✓	✓		✓
5		✓			✓
6			✓		
7	✓				
8					

c



d



effector to a different effector-binding domain (Box 2). In short, redistributing the ensemble can modulate the mechanism. As shown in Box 2, remodelling of the ensemble with effector 2 corresponds to dialing-in (in a rheostat-like fashion) the negative coupling contribution to the overall allosteric coupling.

Discussion

The importance of an ensemble representation of allostery is that it highlights the statistical nature of the allosteric coupling process. The classic notion of an allosteric pathway, through a static picture of the protein, often conveys a deterministic (or at least homogeneous) picture of the signal propagation process, wherein each molecule at any instant is in the same conformation or at least can be equally well represented by the same average conformation. The ensemble model states that, in the most general case, activity is an ensemble-weighted contribution of all species present in solution, regardless of how structurally dissimilar they may be. As such, the important questions shift from identification of the single active conformation, or the allosteric pathway through the protein, to questions that focus on identifying the fraction of molecules that have each domain in the active conformation, and/or the fraction of time a particular set of interactions exist when the functional site (and/or the effector-binding domains) is in its active conformation. An important point critical to characterizing an allosteric ensemble is that, although a given effector molecule may bind to only a small fraction of the states in an ensemble, thus directly stabilizing only those states, stabilization will redistribute the entire ensemble, changing the probability of every state (Box 2). Consequently, every state in the ensemble can be regarded as a potential allosteric state, regardless of whether it binds one or a number of effectors or is functionally

©2014 Macmillan Publishers Limited. All rights reserved

silent. Such silent states, simply as a consequence of being populated, attenuate the impact of effector binding on the activity (Box 2). Indeed, the numerous examples that show complex, stability-dependent and tunable coupling[28,30,86] suggest that the principles outlined here are at play in real allosteric systems.

An important question is how to practically leverage the ensemble view of allostery. The most immediate impact could be refocusing inquiry into allosteric mechanism, particularly as it pertains to IDPs, in which structural methods have limited applicability. The ensemble allosteric model described here and elsewhere[8,61,84] suggests that experimental determination of stabilities and coupling energies represents a viable strategy for determining the relevant states in both disordered and structured ensembles and the relative stabilities of each. Of course, this should not be misconstrued as minimizing the importance of structural characterization of IDPs. Significant progress has been made in modelling disordered ensembles at atomic resolution using various experimental constraints[87–91]. The future challenge will be for these computed ensembles to reproduce the observed highly variable and complex switching behaviour seen in many IDPs.

With regard to structured proteins, the ensemble model of allostery provides insight into important lines of inquiry. Specifically, because the sensitivity of an allosteric protein is maximal at the threshold of a transition (Box 1), many important states are likely to be minutely, albeit measurably, populated under functional conditions (that is, between 1 and 5%). As such, experimental and computational strategies to identify low-probability states are critical. In addition to NMR[55,92–95], fundamental progress has been made in detecting such low-populated states combining other structural biology techniques[51,96], with at least one X-ray crystallography study corroborating the states detected by NMR[51]. The ability to observe alternative conformations masked by the cryogenic temperatures typically used in structural analysis[51,97], and to resolve, using ensemble refinement, multiple states from crystal data collected at room temperature[98], reveal a much richer dynamic picture. Similarly, new data analysis methods applied to electron microscopy images quite expectedly reveal multiple conformational states of macromolecules[99].

Interestingly, the ensemble model also provides a potential framework for interpreting long-time scale molecular dynamics simulations of allosteric proteins[100]. If such simulations can provide the amount of time that specific interactions are made (and the co-occurrence with other interactions), molecular dynamics could be used to map the energy landscape of the allosteric process, possibly even illuminating the multitude of mechanisms pre-encoded in the ensemble. Indeed, a recent long-time molecular dynamics simulation[101] reveals the statistical nature of the interactions associated with allosteric drugs of the muscarinic acetylcholine receptor (M2 receptor), indicating that interactions are not obligatorily associated with ligand-bound or -unbound states. Other computational approaches, such as coarse-grained modelling of the normal modes of proteins, yield predictive cooperative energies even in the absence of structural change[73]. Ongoing and future efforts to integrate computation, structure and even different models of allostery and function[102] are likely to be critical to understanding signalling within large macromolecular complexes[103].

Finally, the ensemble representation of allostery is particularly appealing when considering protein design and evolution. The genetic record clearly shows that nature has evolved and adopted a finite number of functional and binding domains, and has combined them in different ways to make new proteins with different functions and regulatory properties[104,105]. The ensemble representation suggests that by simply manipulating either or both the active/inactive equilibrium for different domains and the coupling between domains, evolution or design of allosteric proteins may be much more straightforward than a detailed inspection of the atomic structure of the protein suggests. Future research directions should include whether allosteric proteins can be designed or evolved without detailed consideration of the structure, and whether positively coupled domains can be transformed to negatively coupled ones, simply by identifying the thermodynamic ground rules for switching[84] and rationally changing the relevant parameters[106–108]. Although the answers to such questions are uncertain, what is certain is that as more dynamic and disordered systems have become amenable to experimental inquiry, the ensemble nature of allostery has become clear.

Received 13 September 2013; accepted 3 January 2014.

1. Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26,** 389–401 (1961).
2. Changeux, J. P. The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb. Symp. Quant. Biol.* **26,** 313–318 (1961).
3. Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12,** 88–118 (1965).
4. Freiburger, L. A. *et al.* Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nature Struct. Mol. Biol.* **18,** 288–294 (2011).
5. Nussinov, R., Tsai, C. J. & Ma, B. The (still) underappreciated role of allostery in the cellular network. *Annu. Rev. Biophys.* **42,** 169–189 (2013).
6. Monod, J. *Chance and Necessity: Essay on the Natural Philosophy of Modern Biology* (Penguin Books, 1977).
7. Fenton, A. W. Allostery: an illustrated definition for the ''second secret of life''. *Trends Biochem. Sci.* **33,** 420–425 (2008).
8. Hilser, V. J., Wrabl, J. O. & Motlagh, H. N. Structural and energetic basis of allostery. *Ann. Rev. Biophys.* **41,** 585–609 (2012).
9. Perutz, M. F. *et al.* Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* **185,** 416–422 (1960).
10. Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228,** 726–734 (1970).
11. Dickerson, R. E. X-ray studies of protein mechanisms. *Annu. Rev. Biophys. Chem.* **41,** 815–842 (1972).
12. Laskowski, R. A., Gerick, F. & Thornton, J. M. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **583,** 1692–1698 (2009).
13. Perutz, M. F., Wilkinson, A. J., Paoli, M. & Dodson, G. G. The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 1–34 (1998).
14. Changeux, J. P. & Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **308,** 1424–1428 (2005).
15. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Bioinf.* **57,** 433–443 (2004).
16. Tzeng, S. R. & Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **21,** 62–67 (2011).
17. Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13,** 748–757 (2003).
18. Smock, R. G. & Gierasch, L. M. Sending signals dynamically. *Science* **324,** 198–203 (2009).
19. Tsai, C. J., del Sol, A. & Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. Biosyst.* **5,** 207–216 (2009).
20. Daily, M. D. & Gray, J. J. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLOS Comput. Biol.* **5,** e1000293 (2009).
21. Swain, J. F. *et al.* Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol. Cell* **26,** 27–39 (2007).
22. Zuiderweg, E. R. *et al.* Allostery in the Hsp70 chaperone proteins. *Top. Curr. Chem.* **328,** 99–153 (2013).
23. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl Acad. Sci. USA* **106,** 18249–18254 (2009).
24. Tzeng, S. R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488,** 236–240 (2012).
25. Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372 (2009).
26. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. Dynamically driven protein allostery. *Nature Struct. Mol. Biol.* **13,** 831–838 (2006).
The first experimental demonstrations of dynamically mediated protein allostery in the CAP using relaxation dispersion NMR and NMR-detected hydrogen exchange.
27. Reichheld, S. E., Yu, Z. & Davidson, A. R. The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc. Natl Acad. Sci. USA* **106,** 22263–22268 (2009).
28. Garcia-Pino, A. *et al.* Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142,** 101–111 (2010).
This article demonstrates how increasing the relative concentration of one ligand can result in conditional cooperativity in an intrinsically disordered protein; in other words, the same protein can initially be an on-switch but then an off-switch at higher concentrations.
29. Sevcsik, E., Trexler, A. J., Dunn, J. M. & Rhoades, E. Allostery in a disordered protein: oxidative modifications to α-synuclein act distally to regulate membrane binding. *J. Am. Chem. Soc.* **133,** 7152–7158 (2011).
30. Ferreon, A. C. M., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498,** 390–394 (2013).
This article directly demonstrates cooperative 'switching' behaviour in an intrinsically disordered protein, via allosteric effects from truncation of the amino acid sequence.
31. Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6,** 306–329 (1963).

©2014 Macmillan Publishers Limited. All rights reserved

32. Koshland, D. E., Nemethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5,** 365–385 (1966).
33. Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **17,** 1295–1307 (2008).
A thoughtful, comprehensive review that synthesizes the 'old' and 'new' views of allostery with experimental and computational case studies from the literature.
34. Koshland, D. E. Enzyme flexibility and enzyme action. *J. Cell. Comp. Physiol.* **54,** 245–258 (1959).
35. Whitley, M. J. & Lee, A. L. Frameworks for understanding long-range intra-protein communication. *Curr. Protein Pept. Sci.* **10,** 116–127 (2009).
36. Changeux, J. P. Allostery and the Monod–Wyman–Changeux model after 50 years. *Ann. Rev. Biophys.* **41,** 103–133 (2012).
37. Eaton, W. A. *et al.* Evolution of allosteric models for hemoglobin. *IUBMB Life* **59,** 586–599 (2007).
38. Eaton, W. A., Henry, E. R., Hofrichter, J. & Mozzarelli, A. Is cooperative oxygen binding by hemoglobin really understood? *Nature Struct. Biol.* **6,** 351–358 (1999).
39. Erman, B. A fast approximate method of identifying paths of allosteric communication in proteins. *Proteins Struct. Funct. Bioinf.* **81,** 1097–1101 (2013).
40. Tang, S. *et al.* Predicting allosteric communication in myosin via a pathway of conserved residues. *J. Mol. Biol.* **373,** 1361–1373 (2007).
41. England, J. L. Allostery in protein domains reflects a balance of steric and hydrophobic effects. *Structure* **19,** 967–975 (2011).
42. VanWart, A. T., Eargle, J., Luthey-Schulten, Z. & Amaro, R. E. Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Comput.* **8,** 2949–2961 (2012).
43. Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286,** 295–299 (1999).
44. Süel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10,** 59–69 (2003).
45. Colombo, M. F., Rau, D. C. & Parsegian, V. A. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science* **256,** 655–659 (1992).
Groundbreaking experimental work demonstrating the large energetic effects of hydration on haemoglobin conformation and thus protein allostery.
46. Elber, R. Simulations of allosteric transitions. *Curr. Opin. Struct. Biol.* **21,** 167–172 (2011).
47. Weinkam, P., Chen, Y. C., Pons, J. & Sali, A. Impact of mutations on the allosteric conformational equilibrium. *J. Mol. Biol.* **425,** 647–661 (2013).
48. Marcos, E., Crehuet, R. & Bahar, I. Changes in dynamics upon oligomerization regulate substrate binding and allostery in amino acid kinase family members. *PLOS Comput. Biol.* **7,** e1002201 (2011).
49. Silva, M. M., Rogers, P. H. & Arnone, A. A third quaternary structure of human hemoglobin A at 1.7-Å resolution. *J. Biol. Chem.* **267,** 17248–17256 (1992).
50. Cooper, A. & Dryden, D. T. F. Allostery without conformational change. *Eur. Biophys. J.* **11,** 103–109 (1984).
The first explicit articulation of dynamic allostery, demonstrating the theoretical relevance of an entirely entropic energetic contribution to biological function.
51. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462,** 669–673 (2009).
52. Lukin, J. A. *et al.* Quaternary structure of hemoglobin in solution. *Proc. Natl Acad. Sci. USA* **100,** 517–520 (2003).
53. Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl Acad. Sci. USA* **110,** 12867–12874 (2013).
54. Wand, A. J. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Biol.* **23,** 75–81 (2013).
55. Manley, G., Rivalta, I. & Loria, J. P. Solution NMR and computational methods for understanding protein allostery. *J. Phys. Chem. B* **117,** 3063–3073 (2013).
56. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry* **45,** 6873–6888 (2006).
57. Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* **43,** 1090–1103 (2011).
58. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18,** 343–384 (2005).
59. Wright, P. E. Intrinsically unstructured proteins: re-assessing the structure-function paradigm. *J. Mol. Biol.* **293,** 321–331 (1999).
60. Tompa, P. Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* **21,** 419–425 (2011).
61. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl Acad. Sci. USA* **104,** 8311–8315 (2007).
The first paper to propose and demonstrate that intrinsic disorder can be used by proteins to mediate allosteric coupling.
62. Luque, I. & Freire, E. Structural parameterization of the binding enthalpy of small ligands. *Proteins* **49,** 181–190 (2002).
63. Li, Z., Raychaudhuri, S. & Wand, A. J. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* **5,** 2647–2650 (1996).
64. Yang, D. & Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263,** 369–382 (1996).
65. Igumenova, T. I., Frederick, K. K. & Wand, A. J. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.* **106,** 1672–1699 (2006).
66. Jarymowycz, V. A. & Stone, M. J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* **106,** 1624–1671 (2006).
67. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448,** 325–329 (2007).
68. Lee, A. L., Kinnear, S. A. & Wand, A. J. Redistribution and loss of side chain entropy upon formation of a calmodulin–peptide complex. *Nature Struct. Biol.* **7,** 72–77 (2000).
69. Palmer, A. G., Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339,** 204–238 (2001).
70. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G. & Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nature Chem. Biol.* **6,** 352–358 (2010).
71. Igumenova, T. I., Lee, A. L. & Wand, A. J. Backbone and side chain dynamics of mutant calmodulin–peptide complexes. *Biochemistry* **44,** 12627–12639 (2005).
72. Laine, O., Streaker, E. D., Nabavi, M., Fenselau, C. C. & Beckett, D. Allosteric signaling in the biotin repressor occurs via local folding coupled to global dampening of protein dynamics. *J. Mol. Biol.* **381,** 89–101 (2008).
73. Rodgers, T. L. *et al.* Modulation of global low-frequency motions underlies allosteric regulation: demonstration in CRP/FNR family transcription factors. *PLoS Biol.* **11,** e1001651 (2013).
74. Schrank, T. P., Bolen, D. W. & Hilser, V. J. Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proc. Natl Acad. Sci. USA* **106,** 16984–16989 (2009).
75. Gao, J. & Xu, D. Correlation between posttranslational modification and intrinsic disorder in protein. *Pac. Symp. Biocomput.* 94–103 (2012).
76. Romero, P. R. *et al.* Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl Acad. Sci. USA* **103,** 8390–8395 (2006).
77. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254,** 1598–1603 (1991).
78. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4,** 10–19 (1997).
79. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48,** 545–600 (1997).
80. Pan, H., Lee, J. C. & Hilser, V. J. Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl Acad. Sci. USA* **97,** 12020–12025 (2000).
81. Bray, D. & Duke, T. A. Conformational spread: the propagation of allosteric states in large multiprotein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **33,** 53–73 (2004).
82. Luque, I., Leavitt, S. A. & Freire, E. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.* **31,** 235–256 (2002).
83. Ackers, G. K., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by the lambda phage repressor. *Proc. Natl Acad. Sci. USA* **79,** 1129–1133 (1982).
84. Motlagh, H. N. & Hilser, V. J. Agonsim/antagonism switching in allosteric ensembles. *Proc. Natl Acad. Sci. USA* **109,** 4134–4139 (2012).
85. Bai, F. *et al.* Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science* **327,** 685–689 (2010).
86. Gekko, K., Obu, N., Li, J. & Lee, J. C. A linear correlation between the energetics of allosteric communication and protein flexibility in the *Escherichia coli* cyclic AMP receptor protein revealed by mutation-induced changes in compressibility and amide hydrogen-deuterium exchange. *Biochemistry* **43,** 3844–3852 (2004).
87. Fisher, C. K. & Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **21,** 426–431 (2011).
88. Forman-Kay, J. D. & Mittag, T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **21,** 1492–1499 (2013).
89. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).
90. Bernadó, P. *et al.* A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl Acad. Sci. USA* **102,** 17002–17007 (2005).
91. Jensen, M. R. *et al.* Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* **17,** 1169–1185 (2009).
92. Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl Acad. Sci. USA* **104,** 9615–9620 (2007).
93. Clore, G. M. Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. *Mol. Biosyst.* **4,** 1058–1069 (2008).
94. Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433,** 128–132 (2005).
95. Tang, C., Louis, J. M., Aniana, A., Suh, J. Y. & Clore, G. M. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* **455,** 693–696 (2008).
96. Yu, B. *et al.* Structural and energetic mechanisms of cooperative autoinhibition and activation of Vav1. *Cell* **140,** 246–256 (2010).

©2014 Macmillan Publishers Limited. All rights reserved

97. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl Acad. Sci. USA* **108,** 16247–16252 (2011).
98. Burnley, T. B., Afonine, P. V., Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1,** e00311 (2012).
99. Penczek, P. A., Kimmel, M. & Spahn, C. M. T. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure* **19,** 1582–1590 (2011).
100. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41,** 429–452 (2012).
101. Dror, R. O. *et al.* Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **503,** 295–299 (2013).
102. Qian, H. Cyclic conformational modification of an enzyme: serial engagement, energy relay, hysteretic enzyme, and Fischer's hypothesis. *J. Phys. Chem. B* **114,** 16105–16111 (2010).
103. Ward, A. B., Sali, A. & Wilson, I. A. Integrative structural biology. *Science* **339,** 913–915 (2013).
104. Andreeva, A. *et al.* Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **36,** D419–D425 (2008).
105. Cesareni, G., Gimona, M., Sudol, M. & Yaffe, M. *Modular Protein Domains* (Wiley-VCH, 2005).
106. Loh, S. N. & Ha, J. H. Protein conformational switches: from nature to design. *Chemistry* **18,** 7984–7999 (2012).
107. Choi, J. H., San, A. & Ostermeier, M. Non-allosteric enzyme switches possess larger effector-induced changes in thermodynamic stability than their non-switch analogs. *Protein Sci.* **22,** 475–485 (2013).
108. Zayner, J. P., Antoniou, C., French, A. R., Hause, R. J. Jr & Sosnick, T. R. Investigating models of protein function and allostery with a widespread mutational analysis of a light activated protein. *Biophys. J.* **105,** 1027–1036 (2013).

Acknowledgements This work was supported by National Science Foundation grant MCB1330211 and by National Institutes of Health grants GM63747 and T32-GM008403.

Author Contributions V.J.H. conceived the manuscript; H.N.M., J.O.W., J.L. and V.J.H. wrote and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.J.H. (Hilser@jhu.edu).

©2014 **Macmillan Publishers Limited. All rights reserved**

ARTICLE

doi:10.1038/nature13231

Enantioselective construction of remote quaternary stereocentres

Tian-Sheng Mei[1], Harshkumar H. Patel[1] & Matthew S. Sigman[1]

Small molecules that contain all-carbon quaternary stereocentres—carbon atoms bonded to four distinct carbon substituents—are found in many secondary metabolites and some pharmaceutical agents. The construction of such compounds in an enantioselective fashion remains a long-standing challenge to synthetic organic chemists. In particular, methods for synthesizing quaternary stereocentres that are remote from other functional groups are underdeveloped. Here we report a catalytic and enantioselective intermolecular Heck-type reaction of trisubstituted-alkenyl alcohols with aryl boronic acids. This method provides direct access to quaternary all-carbon-substituted β-, γ-, δ-, ε- or ζ-aryl carbonyl compounds, because the unsaturation of the alkene is relayed to the alcohol, resulting in the formation of a carbonyl group. The scope of the process also includes incorporation of pre-existing stereocentres along the alkyl chain, which links the alkene and the alcohol, in which the stereocentre is preserved. The method described allows access to diverse molecular building blocks containing an enantiomerically enriched quaternary centre.

The quaternary stereocentre is a common structural motif in many natural products and pharmaceuticals[1–3]. However, the synthesis of these stereocentres in a catalytic and enantioselective manner is a formidable challenge, especially in acyclic systems[4]. Typically, quaternary stereocentres are prepared from substrates with pre-existing functional groups adjacent to the site of reaction, whereas methods to access quaternary stereocentres distant from such groups present a considerable, ongoing synthetic hurdle. The most common enantioselective and catalytic approaches use a carbonyl as a functional handle, wherein α-functionalization, by means of alkylation or aldol reactions[4,5], can be accomplished through the reaction of enolate equivalents[6–11] (**I** in Fig. 1a). Enantioselective β-functionalization of a carbonyl can be accomplished through 1,4-conjugate-addition-type processes using various transition metals and coupling partners[12–16] (**II** in Fig. 1a). A powerful alternative to the carbonyl as a pre-installed functional group is the allylic electrophile[17–19] or nucleophile[20], which yields a quaternary centre adjacent to an alkene[21–23] (**III** in Fig. 1a). However, in all of these approaches, the location of C–C bond formation relative to the functional group is strictly defined, which makes it difficult to install a quaternary chiral centre at more remote sites.

On the basis of our group's recent success in developing asymmetric redox-relay Heck-type reactions of disubstituted alkenyl alcohols[24,25], we surmised that a site-selective and enantioselective transformation of trisubstituted alkenes could address this synthetic limitation (Fig. 1b). By applying the proposed method, it is possible to position the alcohol at different chain lengths from the alkene to obtain a diverse range of functionalized carbonyl products. This is a mechanistic consequence of the process. Specifically, site-selective migratory insertion[26,27] of an alkene into the organometallic intermediate produces a Pd alkyl, **B**, that can migrate towards the alcohol through a sequential β-hydride elimination/migratory-insertion process (Fig. 1b, **D** → **E**) ultimately to release the desired carbonyl product, **C**[28,29]. Although well-known Heck cyclization reactions have been developed and extensively applied to the formation of quaternary centres by intramolecular reaction of trisubstituted alkenes[30–32], no examples of catalytic, asymmetric, quaternary stereocentres synthesized through intermolecular Heck-type reactions of isolated (non-conjugated) trisubstituted alkenes are known[2].

We had several concerns at the outset of this work, including questions regarding reactivity, site selectivity and enantioselectivity when using trisubstituted alkenes in intermolecular Heck-type reactions. Acyclic, non-conjugated trisubstituted alkenes are rare substrates in

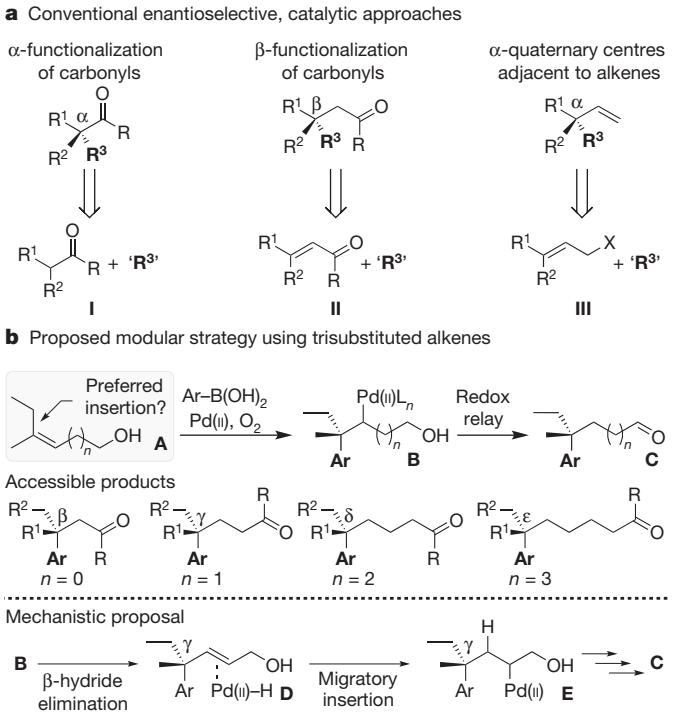


Figure 1 | Approaches to constructing acyclic all-carbon quaternary stereocentres. a, Conventional enantioselective, catalytic approaches: α-functionalization of carbonyls (**I**); β-functionalization of carbonyls (**II**); α-quaternary centres adjacent to alkenes (**III**). **b**, Proposed modular strategy using a redox-relay enantioselective Heck reaction of trisubstituted alkenes and resulting mechanistic analysis. Ar, aryl.

[1]Department of Chemistry, The University of Utah, 315 South, 1400 East, Salt Lake City, Utah 84112, USA.

©2014 Macmillan Publishers Limited. All rights reserved

intermolecular Heck-type reactions, probably because of poor binding to the catalyst or slow migratory insertion[33]. If a reaction does occur, the question of site selectivity is intriguing because the ability to forge a quaternary centre relies on addition to the more substituted carbon. In our study of the redox-relay Heck reaction of disubstituted alkenes, subtle electronic variance of the alkenyl carbons, as determined by ^{13}C chemical shift differences, correlates with site selectivity, with the aryl nucleophile adding to the carbon that is more downfield-shifted[24]. Additional support for electronically influenced site selectivity was revealed by recent density functional theory calculations on this reaction[34,35]. These studies show that site selectivity is controlled by remote dipole interactions of the attached alcohol. Taken together, these observations suggested that, in the case of a trisubstituted alkene, insertion should occur preferentially at the more substituted carbon (the hindered and downfield-shifted carbon). This would possibly also relieve steric strain because the bulky Pd catalyst is positioned at the less hindered carbon. Our final concern was whether this process would be highly enantioselective, given that *cis*- and *trans*-disubstituted alkenes have previously yielded enantiomers as products[24,25]. Considering that trisubstituted alkenes contain both of these stereochemical relationships, the outcome is not easily predicted.

Reaction optimization and scope

We began our investigation by revisiting our previously developed catalytic system[24] for enantioselective oxidative Heck reactions[36–40] of disubstituted alkenes. A trisubstituted homoallylic alcohol (**1**), which has ethyl and methyl groups at the terminus of the alkene, was selected as a model substrate (Fig. 2). Any success with this substrate would bode well for expanding the scope of the reaction to substrates containing other substituents on the alkene with more pronounced differences. Our initial efforts resulted in poor conversion to the desired product, **2a** (40% conversion, 23% yield). Nevertheless, migratory insertion occurred to install the aryl group predominantly at the γ-position (>15 times more likely than at the β-position) and the product was generated in a high enantiomeric ratio (e.r.) of 97:3 (Supplementary Table 1). Encouraged by this initial result, we explored various changes to the reaction conditions, yet these afforded little noticeable improvements in yield. During our previous studies, we observed that the arylboronic acid coupling partner was consumed by various side reactions, such as decomposition of the boronic acid into a phenol and homocoupling of this reagent[41–43]. Indeed, we detected that the arylboronic acid was consumed after 24 h, with corresponding poor conversion of the alkene. We speculated that slow addition of the arylboronic acid would suppress the undesired

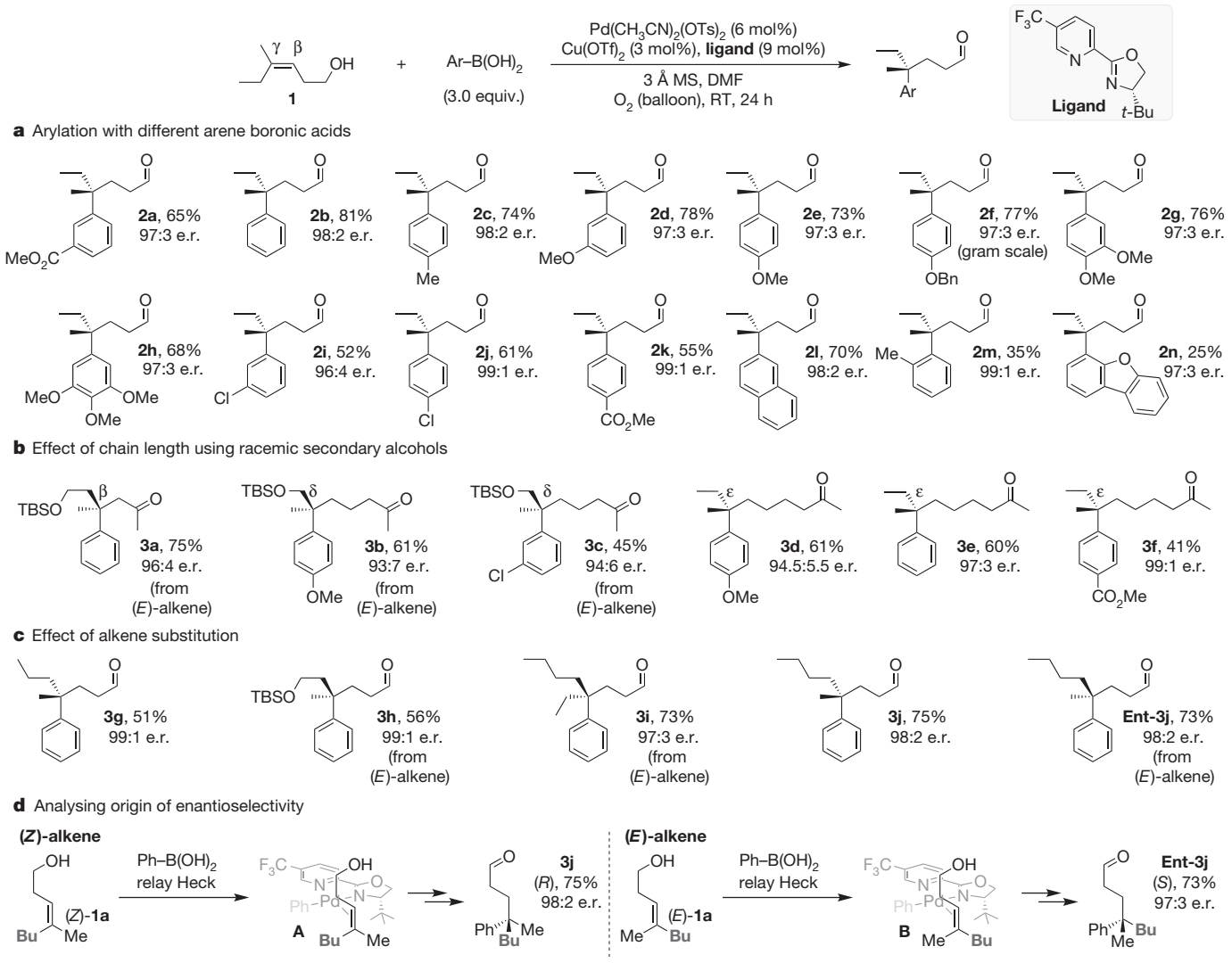


Figure 2 | Enantioselective construction of remote quaternary stereocentres. Conditions for **2a**, **2i**–**2n**, **3c**, **3f**, **3i**–**3j**: 10 mol% $Pd(CH_3CN)_2(OTs)_2$, 4 mol% $Cu(OTf)_2$, 14 mol% ligand, 3 equiv. $ArB(OH)_2$ (two-batch addition, 12 h between additions). *t*-Bu, *t*-butyl; OTs, tosylate; OTf, trifluoromethanesulphonate; MS, molecular sieves; RT, room temperature. Reported yield and e.r. values are an average of at least two experiments. **a**, Exploration of scope of process using various arylboronic acids. Me, methyl. **b**, Evaluation of various chain lengths between the alkene and the alcohol on the substrate. TBS, *t*-butyldimethylsilyl. **c**, Exploration of the alkene substituents. **d**, Proposed origin of enantioselectivity as a function of alkene geometry. Ph, phenyl.

©2014 Macmillan Publishers Limited. All rights reserved

pathways and favour product formation. Batchwise addition of the arylboronic acid did improve the yield to 50%. Increasing the catalyst loading led to 65% yield (**2a** in Fig. 2a), with >15:1 site selectivity (γ/β) and excellent enantioselectivity (e.r. 97:3). A series of control experiments verified the importance of the various reaction components: the absence of either $Cu(OTf)_2$ (ref. 44) or 3 Å molecular sieves[45] substantially reduced the yield, and when the palladium catalyst was excluded, no reaction was observed (Supplementary Table 1). Both of these additives are frequently used in oxidative Pd catalysis to facilitate reoxidation of Pd(0), although their precise role in this transformation is not currently understood.

The scope of arylboronic-acid coupling partners was investigated with homoallylic alcohol **1** (Fig. 2a). A wide array of arylboronic acids were found to be compatible, delivering the corresponding all-carbon quaternary γ-aryl aldehyde products with uniformly high enantioselectivity (e.r. up to 99:1) and in moderate to good yields (**2a**−**2n**). High site selectivity (γ/β ≥ 15:1) is observed with both electron-deficient and electron-rich arylboronic acids. This stands in contrast to our previous reports on enantioselective redox-relay Heck-type reactions of disubstituted alkenes, where only modest site selectivity was achieved for electron-rich aryl boronic acids[24]. This observed difference suggests that the electronic nature of the alkene dictates site selectivity. Higher yields are achieved with electron-rich arylboronic acids, as compared with their electron-poor counterparts (compare **2e** with **2k**), which is consistent with their greater nucleophilicity facilitating migratory insertion of the presumed alkene complex. In all cases, excellent enantioselectivity is observed and the reaction can be scaled to 10 mmol, yielding >2 g as demonstrated by example **2f**. Not surprisingly, lower yields are observed when *ortho*-substituted arenes are used, as illustrated by **2m** and **2n**, although enantioselectivity remains high. The absolute configuration of a derivative of **2f** was determined to be (*R*) by X-ray crystallography (Supplementary Information).

The effect of chain length (the distance from the alcohol to the alkene) using various racemic, trisubstituted alkenyl secondary alcohols was evaluated (Fig. 2b). Of particular note, high site selectivity and enantioselectivity is observed irrespective of the chain length, enabling access to *β*- (**3a**), *δ*- (**3b**, **3c**) and *ε*-quaternary (**3d**−**3f**) functionalized ketone products. Alkenes bearing another oxygen substituent are well tolerated (**3a**−**3c**), although substrates with a styrene unit are unreactive under the present reaction conditions. Substrates were also selected to probe the effect of differential size of the alkene aliphatic substituents (**3g**−**3j** in Fig. 2c). Excellent site selectivity and enantioselectivity is again observed in all cases. To highlight this, an alkene featuring an ethyl group and a butyl group, which have negligible steric differences, performs well, yielding **3i** in 97:3 e.r.

Mechanistic experiments

From analysis of the scope of the reaction, it seems that the enantioselectivity is essentially independent of the steric and electronic nature of both reaction partners, which is atypical in enantioselective reactions. To explore this further, we probed the effect of alkene geometry on enantioselection by comparing the reaction of (*Z*)-**1a** and (*E*)-**1a** (Fig. 2d). The magnitude of the enantioselectivity is the same for both substrates, again suggesting a robust enantioselective reaction. However, the major enantiomer produced in the two cases differ. This is consistent with the binding orientation of the alkene not changing. Specifically,

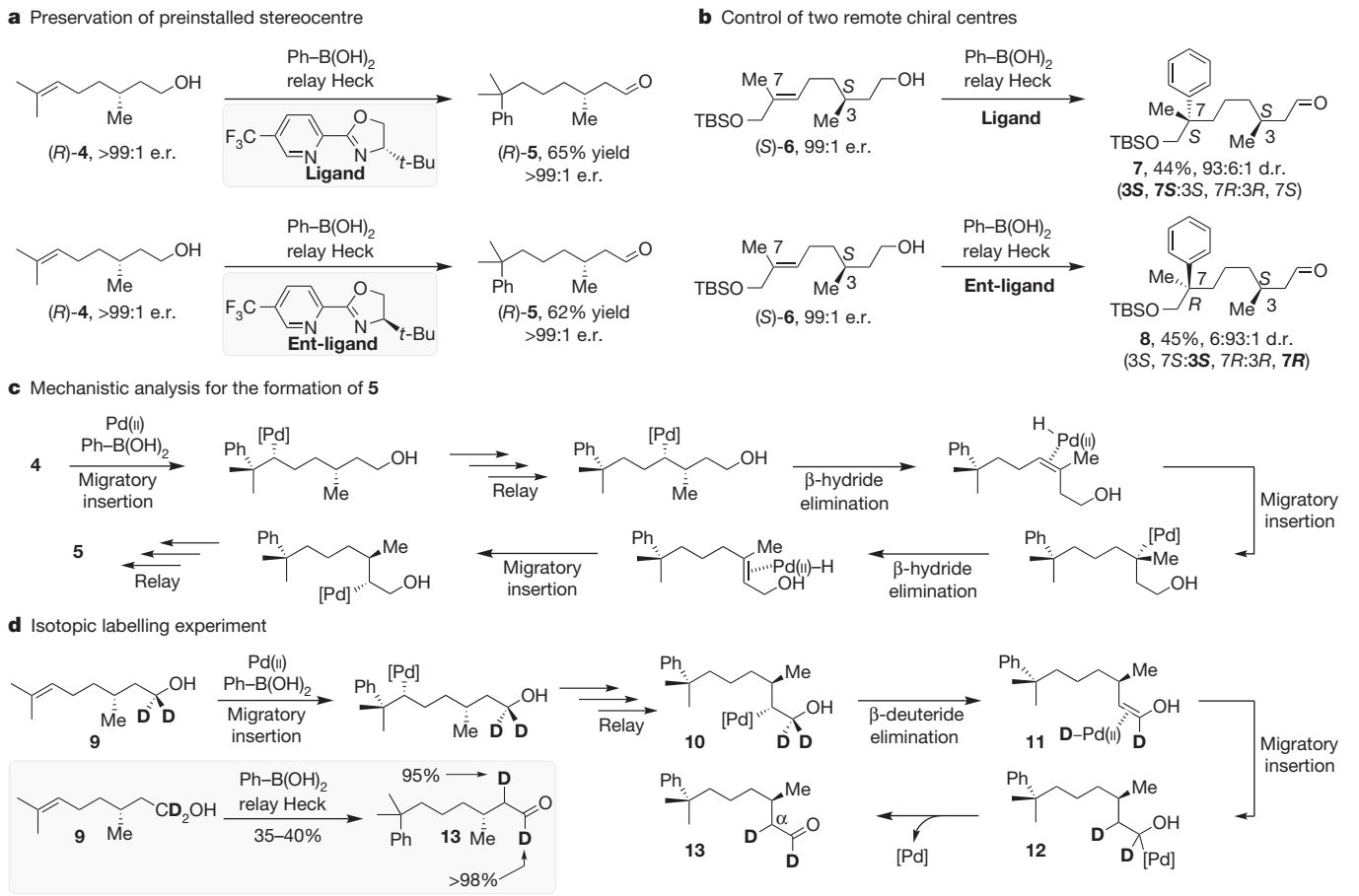


Figure 3 | Evaluation of alkene substrates containing a branch point. Conditions: 10 mol% $Pd(CH_3CN)_2(OTs)_2$, 4 mol% $Cu(OTf)_2$, 14 mol% ligand, 3 equiv. $PhB(OH)_2$. **a**, Independence of catalyst enantiomer on the conservation of the chiral centre during the proposed chain-walking process. **b**, Accessing distinct diastereomers using a combination of catalyst- and substrate-controlled asymmetric synthesis. d.r., diastereomeric ratio. **c**, Proposed mechanistic origin for the observed formation of **5** from **4**. **d**, Isotopic labelling experiment and analysis.

©2014 Macmillan Publishers Limited. All rights reserved

comparison of the proposed intermediates **A** and **B** (Fig. 2d) shows that the alkenyl carbon closer to the alcohol remains fixed, leading to the observed stereochemical outcomes. This conclusion is supported by the relative insensitivity of the process to the group that appears on the terminal end of the trisubstituted alkenes. Although the precise details of why this catalyst is exceptionally selective are under further investigation, these results indicate that few synthetic limitations should be encountered in variation of the alkenyl aliphatic substituents.

As suggested in the initial mechanistic proposal, the Pd catalyst presumably migrates along the alkyl chain until the aldehyde is formed. Indeed, computational studies[34,35] of the redox-relay Heck reaction of disubstituted alkenes shows generally low energy barriers for the 'chain-walking' events[46–48]. Therefore, a key question, with implications for the applicability of this method in more complex settings, is whether the catalyst disengages during the chain-walking process. To explore this possibility, a natural-product-derived substrate, (*R*)-**4**, containing a pre-installed stereogenic centre in the alkyl chain was evaluated using both enantiomers of the catalyst. Preservation of the enantiomeric composition was observed when treating this substrate with either catalyst enantiomer under redox-relay Heck conditions, to yield (*R*)-**5** (Fig. 3a). This implies that as the catalyst proceeds through the iterative β-hydride elimination/migratory-insertion events depicted in Fig. 3c, the catalyst remains both ligated to the substrate and on the same face of the alkene throughout the relay process. As a more striking example, alkene (*S*)-**6** was treated with both enantiomers of catalyst to yield the relay products **7** and **8** in high diastereoselectivity (Fig. 3b). Two distinct diastereomers are produced by the use of different enantiomers of catalyst, because the initial migratory insertion is under catalyst-controlled face selection, but the preset stereogenic centre is not altered during the relay process.

To support the chain walking proposal further, we carried out an isotopic labelling experiment (Fig. 3d). A deuterium-labelled analogue of **4**, alkenol **9**, bearing deuterium atoms at the carbon connecting to the alcohol, was synthesized and submitted to the redox-relay Heck reaction. The experiment reveals clean repositioning of one deuterium atom at the site α to the carbonyl group in the product (**13**) (Fig. 3d). This result is consistent with a mechanism whereby the Pd catalyst migrates through the chain to form intermediate **10**, which undergoes *β*-deuteride elimination followed by reinsertion to yield intermediate **12** (Fig. 3d).

Conclusion

We have described a catalytic and enantioselective addition of arylboronic acids to trisubstituted alkenes that is highly site selective for the more hindered position. The method does not rely on a defined relationship between the site of addition and an adjacent functional group, and thus provides a modular method to access quaternary stereocentres in high enantioselectivity. Furthermore, we anticipate that the mechanistic implications of both site-selective addition of an organometallic to a trisubstituted alkene and the ability of the catalyst to migrate through existing chiral centres will inspire further studies in this area.

METHODS SUMMARY

To a dry, 100-ml Schlenk flask equipped with a stir bar, we added $Pd(CH_3CN)_2(OTs)_2$ (15.9 mg, 0.030 mmol, 6.0 mol%), $Cu(OTf)_2$ (5.4 mg, 0.015 mmol, 3.0 mol%), ligand (12.3 mg, 0.045 mmol, 9.0 mol%), 3 Å molecular sieves (75.0 mg, 150 mg $mmol^{-1}$) and dimethylformamide (DMF; 8 ml). To this flask, a three-way adaptor fitted with a balloon of O_2 was added, and the flask was evacuated via house vacuum and refilled with O_2 three times while stirring. The resulting mixture was stirred for 10 min. To this, a DMF solution (2 ml) of the alkenyl alcohol (0.5 mmol) and corresponding boronic acid (1.5 mmol, 3 equiv.) was added by syringe. The resulting mixture was stirred for 24 h at room temperature (23–25 °C). The mixture was diluted with diethyl ether (200 ml) and water (50 ml). The aqueous layer was extracted with diethyl ether (2 × 50 ml). The combined organic layers were washed with water (3 × 20 ml) and brine (20 ml) and dried over sodium sulphate. The organic extracts were concentrated under reduced pressure, and the resulting residue was purified by silica gel flash chromatography using 2–10% ethyl acetate in hexanes containing 0.1% triethylamine to yield an aldehyde product. Full experimental details and characterization of new compounds can be found in the Supplementary Information.

Received 22 November 2013; accepted 4 March 2014.
Published online 9 April 2014.

1. Trost, B. M. & Jiang, C. Catalytic enantioselective construction of all-carbon quaternary stereocenters. *Synthesis* 369–396 (2006).
2. Douglas, C. J. & Overman, L. E. Catalytic asymmetric synthesis of all-carbon quaternary stereocenters. *Proc. Natl Acad. Sci. USA* **101,** 5363–5367 (2004).
3. Christoffers, J. & Mann, A. Enantioselective construction of quaternary stereocenters. *Angew. Chem. Int. Ed.* **40,** 4591–4597 (2001).
4. Das, J. P. & Marek, I. Enantioselective synthesis of all-carbon quaternary stereogenic centers in acyclic systems. *Chem. Commun.* **47,** 4593–4623 (2011).
5. Marek, I. *et al.* All-carbon quaternary stereogenic centers in acyclic systems through the creation of several C–C bonds per chemical step. *J. Am. Chem. Soc.* **136,** 2682–2694 (2014).
6. Masarwa, A. *et al.* Merging allylic carbon-hydrogen and selective carbon-carbon bond activation. *Nature* **505,** 199–203 (2014).
7. Liu, W.-B., Reeves, C. M. & Stoltz, B. M. Enantio-, diastereo-, and regioselective iridium-catalyzed asymmetric allylic alkylation of acyclic β-ketoesters. *J. Am. Chem. Soc.* **135,** 17298–17301 (2013).
8. Krautwald, S., Sarlah, D., Schafroth, M. A. & Carreira, E. M. Enantio- and diastereodivergent dual catalysis: α-allylation of branched aldehydes. *Science* **340,** 1065–1068 (2013).
9. Taylor, M. S. & Jacobsen, E. N. Enantioselective Michael additions to α,β-unsaturated imides catalyzed by a salen-Al complex. *J. Am. Chem. Soc.* **125,** 11204–11205 (2003).
10. Mermerian, A. H. & Fu, G. C. Catalytic enantioselective synthesis of quaternary stereocenters via intermolecular C-acylation of silyl ketene acetals: dual activation of the electrophile and the nucleophile. *J. Am. Chem. Soc.* **125,** 4050–4051 (2003).
11. Sawamura, M., Hamashima, H. & Ito, Y. Catalytic asymmetric synthesis with trans-chelating chiral diphosphine ligand TRAP: rhodium-catalyzed asymmetric Michael addition of α-cyano carboxylates. *J. Am. Chem. Soc.* **114,** 8295–8296 (1992).
12. Mauleón, P. & Carretero, J. C. Enantioselective construction of stereogenic quaternary centers via Rh-catalyzed asymmetric addition of alkenylboronic acids to α,β-unsaturated pyridylsulfones. *Chem. Commun.* 4961–4963 (2005).
13. Wu, J., Mampreian, D. M. & Hoveyda, A. H. Enantioselective synthesis of nitroalkanes bearing all-carbon quaternary stereogenic centers through Cu-catalyzed asymmetric conjugate additions. *J. Am. Chem. Soc.* **127,** 4584–4585 (2005).
14. Hawner, C. *et al.* Rhodium-catalyzed asymmetric 1,4-addition of aryl alanes to trisubstituted enones: BINAP as an effective ligand in the formation of quaternary stereocenters. *Angew. Chem. Int. Ed.* **49,** 7769–7772 (2010).
15. Shintani, R., Takeda, M., Nishimura, T. & Hayashi, T. Chiral tetrafluorobenzobarrelenes as effective ligands for rhodium-catalyzed asymmetric 1,4-addition of arylboroxines to β,β-disubstituted α,β-unsaturated ketones. *Angew. Chem. Int. Ed.* **49,** 3969–3971 (2010).
16. Dabrowski, J. A., Villaume, M. T. & Hoveyda, A. H. Enantioselective synthesis of quaternary carbon stereogenic centers through copper-catalyzed conjugate additions of aryl- and alkylaluminum reagents to acyclic trisubstituted enones. *Angew. Chem. Int. Ed.* **52,** 8156–8159 (2013).
17. Jung, B. & Hoveyda, A. H. Site- and enantioselective formation of allene-bearing tertiary or quaternary carbon stereogenic centers through NHC–Cu-catalyzed allylic substitution. *J. Am. Chem. Soc.* **134,** 1490–1493 (2012).
18. Lee, Y. & Hoveyda, A. H. Lewis base activation of Grignard reagents with N-heterocyclic carbenes. Cu-free catalytic enantioselective additions to γ-chloro-α,β-unsaturated esters. *J. Am. Chem. Soc.* **128,** 15604–15605 (2006).
19. Luchaco-Cullis, C. A., Mizutani, H., Murphy, K. E. & Hoveyda, A. H. Modular pyridinyl peptide ligands in asymmetric catalysis: enantioselective synthesis of quaternary carbon atoms through copper-catalyzed allylic substitutions. *Angew. Chem. Int. Ed.* **40,** 1456–1460 (2001).
20. Denmark, S. E. & Fu, J. Catalytic, enantioselective addition of substituted allylic trichlorosilanes using a rationally-designed 2,2′-bispyrrolidine-based bisphosphoramide. *J. Am. Chem. Soc.* **123,** 9488–9489 (2001).
21. Shi, W.-J. *et al.* Highly enantioselective hydrovinylation of α-alkyl vinylarenes. An approach to the construction of all-carbon quaternary stereocenters. *J. Am. Chem. Soc.* **128,** 2780–2781 (2006).
22. Zhang, A., &. RajanBabu, T. V. All-carbon quaternary centers via catalytic asymmetric hydrovinylation. New approaches to the exocyclic side chain stereochemistry problem. *J. Am. Chem. Soc.* **128,** 5620–5621 (2006).
23. Zhang, P., Le, H., Kyne, R. E. & Morken, J. P. Enantioselective construction of all-carbon quaternary centers by branch-selective Pd-catalyzed allyl-allyl cross-coupling. *J. Am. Chem. Soc.* **133,** 9716–9719 (2011).
24. Mei, T.-S., Werner, E. W., Burckle, A. J. & Sigman, M. S. Enantioselective redox-relay oxidative Heck arylations of acyclic alkenyl alcohols using boronic acids. *J. Am. Chem. Soc.* **135,** 6830–6833 (2013).
25. Werner, E. W., Mei, T.-S., Burckle, A. J. & Sigman, M. S. Enantioselective Heck arylations of acyclic alkenyl alcohols using a redox-relay strategy. *Science* **338,** 1455–1458 (2012).
26. Sigman, M. S. & Werner, E. W. Imparting catalyst control upon classical palladium-catalyzed alkenyl C–H bond functionalization reactions. *Acc. Chem. Res.* **45,** 874–884 (2012).
27. Beletskaya, I. P. & Cheprakov, A. V. The Heck reaction as a sharpening stone of palladium catalysis. *Chem. Rev.* **100,** 3009–3066 (2000).

©2014 Macmillan Publishers Limited. All rights reserved

28. Calò, V., Nacci, A., Monopoli, A. & Ferola, V. Palladium-catalyzed Heck arylations of allyl alcohols in ionic liquids: remarkable base effect on the selectivity. *J. Org. Chem.* **72,** 2596–2601 (2007).
29. Bouquillon, S., Ganchegui, B., Estrine, B., Hénin, F. & Muzart, J. Heck arylation of allylic alcohols in molten salts. *J. Organomet. Chem.* **634,** 153–156 (2001).
30. Dounay, A. B. & Overman, L. E. The asymmetric intramolecular Heck reaction in natural product total synthesis. *Chem. Rev.* **103,** 2945–2964 (2003).
31. Shibasaki, M., Vogl, E. M. & Ohshima, T. Asymmetric Heck reaction. *Adv. Synth. Catal.* **346,** 1533–1552 (2004).
32. Mc Cartney, D. & Guiry, P. J. The asymmetric Heck and related reactions. *Chem. Soc. Rev.* **40,** 5122–5150 (2011).
33. Melpolder, J. B. & Heck, R. F. Palladium-catalyzed arylation of allylic alcohols with aryl halides. *J. Org. Chem.* **41,** 265–272 (1976).
34. Dang, Y., Qu, S., Wang, Z.-X. & Wang, X. A Computational mechanistic study of an unprecedented Heck-type relay reaction: insight into the origins of regio- and enantioselectivities. *J. Am. Chem. Soc.* **136,** 986–998 (2014).
35. Xu, L. *et al.* Mechanism, reactivity, and selectivity in palladium-catalyzed redox-relay Heck arylations of alkenyl alcohols. *J. Am. Chem. Soc.* **136,** 1960–1967 (2014).
36. Oliveira, C. C., Angnes, R. A. & Correia, C. R. D. Intermolecular enantioselective Heck–Matsuda arylations of acyclic olefins: application to the synthesis of β-aryl-γ-lactones and β-aryl aldehydes. *J. Org. Chem.* **78,** 4373–4385 (2013).
37. Kikushima, K., Holder, J. C., Gatti, M. & Stoltz, B. M. Palladium-catalyzed asymmetric conjugate addition of arylboronic acids to five-, six-, and seven-membered β-substituted cyclic enones: enantioselective construction of all-carbon quaternary stereocenters. *J. Am. Chem. Soc.* **133,** 6902–6905 (2011).
38. Yoo, K. S. *et al.* Asymmetric intermolecular boron Heck-type reactions via oxidative palladium(II) catalysis with chiral tridentate NHC-amidate-alkoxide ligands. *J. Org. Chem.* **75,** 95–101 (2010).
39. Yoo, K. S. *et al.* Asymmetric intermolecular Heck-type reaction of acyclic alkenes via oxidative palladium(II) catalysis. *Org. Lett.* **9,** 3933–3935 (2007).
40. Yonehara, K. *et al.* Palladium-catalyzed asymmetric intermolecular arylation of cyclic or acyclic alkenes using phosphinite-oxazoline ligands derived from d-glucosamine. *J. Organomet. Chem.* **603,** 40–49 (2000).
41. Werner, E. W. & Sigman, M. S. A highly selective and general palladium catalyst for the oxidative Heck reaction of electronically nonbiased olefins. *J. Am. Chem. Soc.* **132,** 13981–13983 (2010).
42. Yoo, K. S., Yoon, C. H. & Jung, K. W. Oxidative palladium(II) catalysis: a highly efficient and chemoselective cross-coupling method for carbon−carbon bond formation under base-free and nitrogenous-ligand conditions. *J. Am. Chem. Soc.* **128,** 16384–16393 (2006).
43. Du, X. *et al.* Mizoroki−Heck type reaction of organoboron reagents with alkenes and alkynes. A Pd(II)-catalyzed pathway with $Cu(OAc)_2$ as an oxidant. *Org. Lett.* **3,** 3313–3316 (2001).
44. Gligorich, K. M. & Sigman, M. S. Recent advancements and challenges of palladium^{II}-catalyzed oxidation reactions with molecular oxygen as the sole oxidant. *Chem. Commun.* 3854–3867 (2009).
45. Steinhoff, B. A., King, A. E. & Stahl, S. S. Unexpected roles of molecular sieves in palladium-catalyzed aerobic alcohol oxidation. *J. Org. Chem.* **71,** 1861–1868 (2006).
46. Kochi, T., Hamasaki, T., Aoyama, Y., Kawasaki, J. & Kakiuchi, F. Chain-walking strategy for organic synthesis: catalytic cycloisomerization of 1,*N*-dienes. *J. Am. Chem. Soc.* **134,** 16544–16547 (2012).
47. Johnson, L. K., Killian, C. M. & Brookhart, M. New Pd(II)- and Ni(II)-based catalysts for polymerization of ethylene and α-olefins. *J. Am. Chem. Soc.* **117,** 6414–6415 (1995).
48. Larock, R. C., Leung, W.-Y. & Stolz-Dunn, S. Synthesis of aryl-substituted aldehydes and ketones via palladium-catalyzed coupling of aryl halides and non-allylic unsaturated alcohols. *Tetrahedr. Lett.* **30,** 6629–6632 (1989).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the US National Institutes of Health (NIGMS GM063540) for financial support.

Author Contributions T.-S.M. and H.H.P. performed the experiments and analysed the data. T.-S.M. and M.S.S. designed the experiments. T.-S.M. and M.S.S. prepared this manuscript with feedback from H.H.P.

Author Information Data for the crystallized product (a derivative of **2f**) have been deposited in the Cambridge Crystallographic Data Centre under accession number CCDC 988090. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.S. (sigman@chem.utah.edu).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13200

Domains of genome-wide gene expression dysregulation in Down's syndrome

Audrey Letourneau[1]*, Federico A. Santoni[1]*, Ximena Bonilla[1], M. Reza Sailani[1], David Gonzalez[2], Jop Kind[3], Claire Chevalier[4], Robert Thurman[5], Richard S. Sandstrom[5], Youssef Hibaoui[6], Marco Garieri[1], Konstantin Popadin[1], Emilie Falconnet[1], Maryline Gagnebin[1], Corinne Gehrig[1], Anne Vannier[1], Michel Guipponi[1], Laurent Farinelli[7], Daniel Robyr[1], Eugenia Migliavacca[1,8], Christelle Borel[1], Samuel Deutsch[9], Anis Feki[6], John A. Stamatoyannopoulos[5], Yann Herault[4], Bas van Steensel[3], Roderic Guigo[2] & Stylianos E. Antonarakis[1,10]

Trisomy 21 is the most frequent genetic cause of cognitive impairment. To assess the perturbations of gene expression in trisomy 21, and to eliminate the noise of genomic variability, we studied the transcriptome of fetal fibroblasts from a pair of monozygotic twins discordant for trisomy 21. Here we show that the differential expression between the twins is organized in domains along all chromosomes that are either upregulated or downregulated. These gene expression dysregulation domains (GEDDs) can be defined by the expression level of their gene content, and are well conserved in induced pluripotent stem cells derived from the twins' fibroblasts. Comparison of the transcriptome of the Ts65Dn mouse model of Down's syndrome and normal littermate mouse fibroblasts also showed GEDDs along the mouse chromosomes that were syntenic in human. The GEDDs correlate with the lamina-associated (LADs) and replication domains of mammalian cells. The overall position of LADs was not altered in trisomic cells; however, the H3K4me3 profile of the trisomic fibroblasts was modified and accurately followed the GEDD pattern. These results indicate that the nuclear compartments of trisomic cells undergo modifications of the chromatin environment influencing the overall transcriptome, and that GEDDs may therefore contribute to some trisomy 21 phenotypes.

Down's syndrome results from total or partial trisomy of chromosome 21. It is the most frequent live-born aneuploidy affecting 1 in 750 infants. Down's syndrome patients are characterized by a cognitive impairment as well as muscle hypotonia, dysmorphic features, Alzheimer's disease neuropathology or congenital heart defects[1]. The severity and the incidence of those phenotypes are variable within the Down's syndrome population[1]. Among the possible causes, the genetic (or epigenetic) background of each individual may contribute to this phenotypic variability.

It is likely that most of the Down's syndrome phenotypes are related to alteration of gene expression due to the supernumerary copy of chromosome 21 (HSA21). According to the 'gene dosage effect' hypothesis, some Down's syndrome features could be directly explained by the dosage imbalance of genes on HSA21 (refs 2–4). Additionally, the phenotypes may also be due to the presence of extra DNA material[3,5]. Understanding the genomic determinants that contribute to the different phenotypes is a major objective in Down's syndrome research[6].

Several Down's syndrome mouse models have been created to mimic the gene expression changes observed in humans. The models are based either on translocation or on duplication of syntenic regions between HSA21 and segments of mouse chromosomes (MMUs) 10, 16 and 17. Among them, the Ts65Dn mouse model has been extensively used to study the molecular mechanisms of the features of Down's syndrome[7]. Ts65Dn mice harbour a translocation of MMU16 and exhibit some of the Down's syndrome phenotypes[8]. Other mouse models are based on the triplication of single candidate genes such as *SIM2* or *DYRK1A*. These models have emphasized the role of individual HSA21 genes in specific phenotypes, in particular the cognitive impairment[9,10].

Because Down's syndrome is probably due to gene expression disturbances, the investigation of the molecular mechanisms that underlie the phenotypic consequences requires an understanding of the transcriptome differences in trisomic cells and tissues. Several studies have explored the changes of gene expression between trisomic individuals and controls[11–14]. However, the extensive natural gene expression variation occurring in both normal and Down's syndrome individuals[15,16] complicates the identification of changes related to trisomy 21 per se. To assess the perturbations of gene expression in Down's syndrome without genetic variation among the samples, we studied a pair of monozygotic twins discordant for trisomy 21 (ref. 17). The use of these samples eliminates the bias of genome variability and thus most of the transcriptome differences observed are probably related to the supernumerary HSA21. Notably, we have found that the differential gene expression between the trisomy 21 discordant twins is organized in domains along all the chromosomes. We show that those domains are conserved in the Ts65Dn mouse model and correlate with the previously described lamina-associated and replication time domains. The study of histone mark profiles confirmed the role of chromatin modifications in the gene expression changes

[1]Department of Genetic Medicine and Development, University of Geneva Medical School, University Hospitals of Geneva, 1211 Geneva, Switzerland. [2]Center for Genomic Regulation, University Pompeu Fabra, 08003 Barcelona, Spain. [3]Division of Gene Regulation, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. [4]AneuPath 21, Institut de Génétique Biologie Moléculaire et Cellulaire, Translational medicine and Neuroscience program, IGBMC, ICS, PHENOMIN, CNRS, INSERM, Université de Strasbourg, UMR7104, UMR964, 1 rue Laurent Fries, 67404 Illkirch, France. [5]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [6]Stem Cell Research Laboratory, Department of Obstetrics and Gynecology, Geneva University Hospitals, 1211 Geneva, Switzerland. [7]FASTERIS SA, 1228 Plan-les-Ouates, Switzerland. [8]Swiss Institute of Bioinfomatics, 1211 Geneva, Switzerland. [9]DOE Joint Genome Institute, Walnut Creek, California 94598, USA. [10]iGE3 Institute of Genetics and Genomics of Geneva, 1211 Geneva, Switzerland.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

observed between the twins. Altogether, our results suggest a new molecular mechanism in the regulation of gene expression in trisomic cells.

Differential expression is organized in domains

We used messenger RNA sequencing to study the transcriptome of fetal skin primary fibroblasts derived from both the trisomic (T1DS) and the normal (T2N) twin, in four replicates for each. In total, 63–157 million 100-base-pair (bp) paired-end reads were generated from each sample and mapped with GEM[18]. The normalized gene expression (RPKM, reads per kilobase per million) was compared between the twins. We used EdgeR[19] to evaluate the differences of gene expression and found 182 genes (including 42 long non-coding RNAs (lncRNAs)) significantly differentially expressed between the twins (false discovery rate (FDR) <0.05) (Supplementary Table 1). The gene ontology analysis[20] revealed a reduced expression of secreted proteins involved in signalling (adjusted $P = 1.2 \times 10^{-8}$) and in particular those involved in cytokine–cytokine receptor interaction pathways (adjusted $P = 8 \times 10^{-4}$) and inflammatory response (adjusted $P = 3.8 \times 10^{-4}$). This confirms that trisomic cells may have a defective cell signalling system contributing to the impairment of the immune system[21].

We then focused on general transcriptomic changes occurring in trisomic cells. We assessed the genome-wide differential expression between the discordant twins by looking at the distribution of gene expression fold changes along the chromosomes. This comparison revealed well-defined chromosomal domains composed of neighbouring genes sharing differential expression profiles. In most of the chromosomes, large regions of upregulated genes alternate with large downregulated domains (Figs 1a, b). This observation suggests that the differential expression between T1DS and T2N is not randomly organized but follows a specific pattern along the chromosomes. We used a smoothing function to define the domain borders and identified a total of 337 GEDDs in the trisomy 21 discordant twins (Supplementary Table 2). Those domains vary in size (from 9 kilobases (kb) to 114 megabases (Mb), median size of 3.2 Mb) and contain up to 507 genes with a median number of 20.

Three independent replicate experiments confirmed these GEDDs in the discordant twins (overall correlation (Spearman correlation) ρ(Rep0, Rep1) = 0.88; ρ(Rep0, Rep2) = 0.76; ρ(Rep0, Rep3) = 0.99, Supplementary Fig. 2). We also compared the gene expression profiles of fibroblasts derived from a healthy pair of monozygotic twins (MZ1 and MZ2) as a control (overall ρ(Rep0, N) = −0.01, Supplementary Fig. 2). The absence of domains in this latter comparison suggests that the domain organization observed in the discordant twins can be mainly attributed to the supernumerary HSA21.

Gene expression levels and GEDDs

Next, we investigated the expression level of the genes within the domains. We classified genes according to low ($0.1 < \text{RPKM} < 10$), medium ($10 < \text{RPKM} < 100$) or high ($\text{RPKM} > 100$) expression level and their position along the chromosomes (Fig. 2a and Supplementary Fig. 3). For each category, we compared the expression fold change distribution with the expected distribution given by the healthy monozygotic twins (Fig. 2b). We found that the highly expressed genes contribute substantially to the domains downregulated in T1DS ($P < 2.2 \times 10^{-16}$ and $P = 1.4 \times 10^{-15}$ for medium and high expression, respectively). In contrast, the low expressed genes are mainly associated with the upregulated regions ($P = 4.4 \times 10^{-4}$). These data show that the expression fold change between trisomic and normal cells is organized in chromosomal domains and that those domains can be partially defined by the expression level of their gene content.

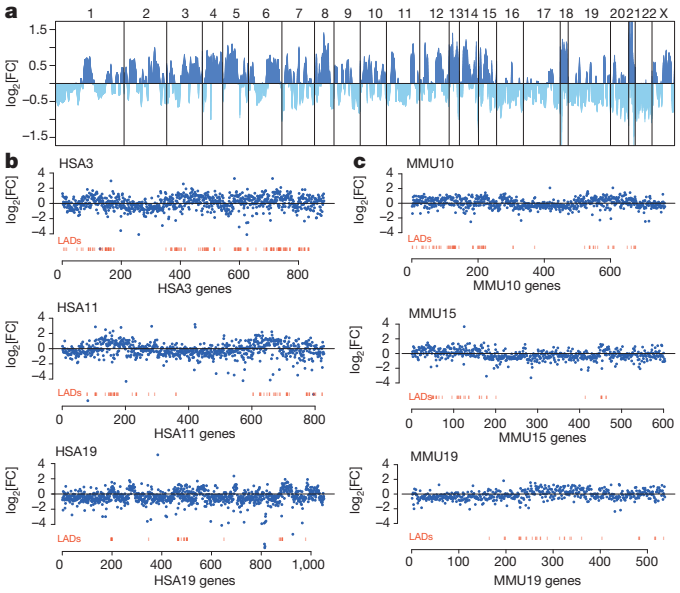


Figure 1 | Gene expression fold change is organized in chromosomal domains. **a**, Lowess smoothed $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along the human genome. Numbers indicate the chromosomes delimited with vertical lines. Upregulated domains are shown in dark blue and downregulated domains in light blue. **b**, $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along human chromosomes 3, 11 and 19. **c**, $\log_2$ fold change of gene expression between Ts65Dn and wild-type littermates along mouse chromosomes 10, 15 and 19. Genes (in blue) are sorted according to their position on the chromosome, at equidistance. The human and mouse LADs[27,28] are shown in red.

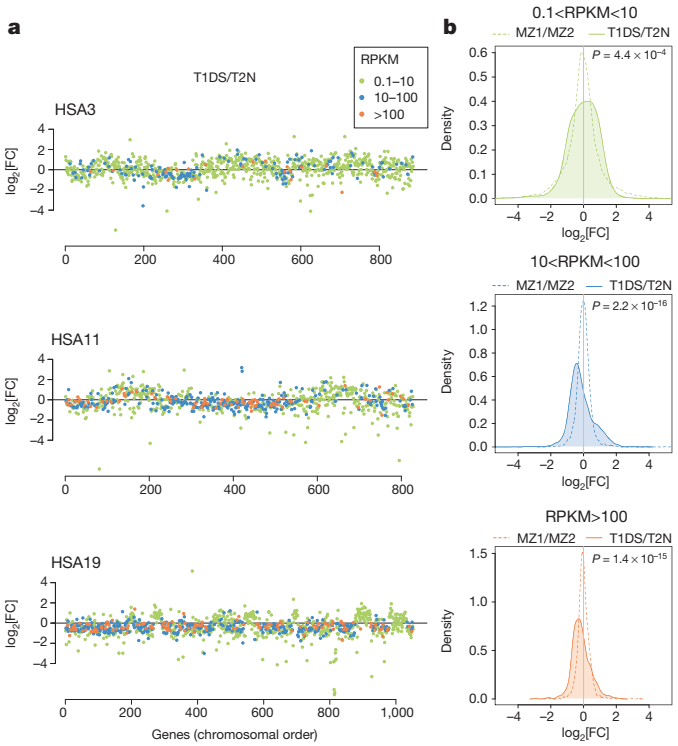


Figure 2 | Weakly and highly expressed genes contribute differently to the domains. **a**, $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along human chromosomes 3, 11 and 19. Genes are sorted according to their position on the chromosome (at equidistance) and coloured according to their expression level. **b**, Distribution of expression fold changes ($\log_2$) for low ($0.1 < \text{RPKM} < 10$, upper panel), medium ($10 < \text{RPKM} < 100$, middle panel) and high ($\text{RPKM} > 100$, lower panel) expressed genes. Solid lines represent the $\log_2$[FC] between the discordant twins (T1DS/T2N Rep0) and dashed lines represent the $\log_2$[FC] between the healthy twins (MZ1/MZ2). P = Wilcoxon test P value.

©2014 Macmillan Publishers Limited. All rights reserved

Previous studies have reported the clustering of highly expressed genes in domains[22,23]. We examined the gene expression along the chromosomes for each sample to understand the respective contribution of each twin to the GEDD pattern. We confirmed the clustering of genes according to their expression levels in both trisomic and normal fibroblasts. However, we found that the trisomic cells show a decreased dynamics of variation between highly and weakly expressed genes in all chromosomes but HSA13 and HSA18 (Supplementary Fig. 4). This difference of gene expression amplitude between trisomic and normal cells contributes to the observed GEDDs and indicates that gene expression might be less fine-tuned and less dynamic in a trisomic context.

GEDDs are conserved in induced pluripotent stem cells

To verify whether the GEDDs described above could be observed in other cell types, we derived induced pluripotent stem (iPS) cells from each of the fibroblast lines from the discordant twins[24]. We performed mRNA sequencing on both iPS lines and compared the normalized gene expression values along the chromosomes. Notably, for most of the chromosomes, we found that the differential expression patterns in iPS cells are highly similar (overall $\rho = 0.85$) to those observed in the fibroblasts (Fig. 3 and Supplementary Fig. 5). These results indicate that the GEDDs are conserved after dedifferentiation and that the supernumerary HSA21 has similar effects in the genome-wide dysregulation of gene expression in fibroblasts and iPS cells.

GEDDs are conserved in mouse Ts65Dn fibroblasts

To verify whether the GEDDs described in the twins' fibroblasts are conserved in other organisms, we performed a similar analysis in the partial trisomy 16 Ts65Dn mouse model of Down's syndrome. We analysed the differential expression pattern between primary cultures of Ts65Dn and normal littermate mouse fibroblasts. We first confirmed the expected increased expression of the MMU16 telomeric region that is triplicated in the Ts65Dn mice (Supplementary Fig. 6a). Moreover, those mice present an additional partial trisomy of the centromeric region of MMU17 that is not syntenic to HSA21 (ref. 25); genes in that region are also overexpressed (Supplementary Fig. 6b). Then we investigated the expression fold change between trisomic and wild-type mice and found that it is organized in GEDDs along the Ts65Dn mouse chromosomes (Fig. 1c and Supplementary Fig. 7). The comparison of the largest syntenic blocks (>1 Mb) between mouse and human revealed that the GEDDs are well conserved between the discordant twins' fibroblasts and the mouse fibroblasts and that the direction of dysregulation is maintained between the two species (Fig. 4a). This conservation is independent of the karyotypic context, as the syntenic regions are located in different chromosomes in the mouse. Human and mouse orthologous genes show similar expression fold changes (overall $\rho = 0.44$) (Fig. 4b, top panel). The same comparison using the healthy monozygotic twins did not show a correlation between human and mouse fold changes (overall $\rho = -0.13$) (Fig. 4b, bottom panel). Those results demonstrate that the GEDDs observed between T1DS and T2N are remarkably conserved in the Ts65Dn mouse model for Down's syndrome and indicate that the domain organization is independent of the chromosomal context.

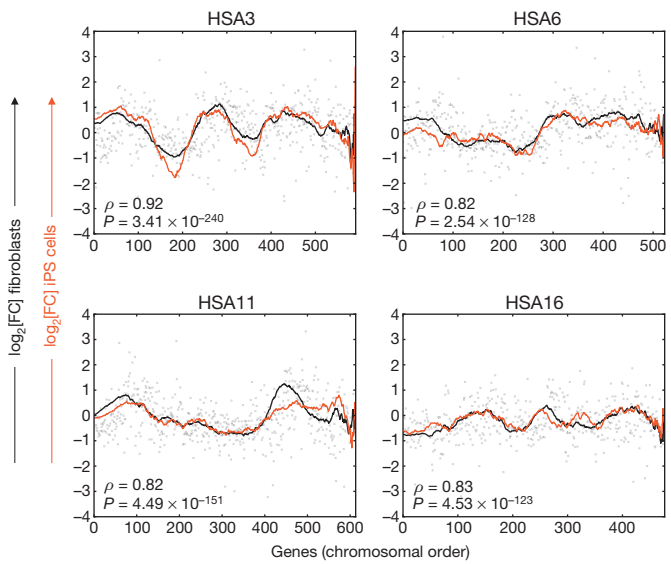


Figure 3 | GEDDs are conserved in iPS cells. Comparison of the $\log_2$ expression fold change profiles between T1DS and T2N in fibroblasts (in black, Rep0) and iPS cells (in red) along human chromosomes 3, 6, 11 and 16. ρ = Spearman correlation; P = associated P value.

Comparison of unrelated individuals

To verify whether GEDDs could be identified when comparing trisomy 21 to normal samples from unrelated individuals, we sequenced the mRNA from fibroblasts of 8 trisomic and 8 sex-matched controls. We reasoned that the natural genetic variability would have masked the domains[26]. Therefore, we averaged the expression values within each group to reduce the influence of individual genetic backgrounds and calculated the expression fold change between the two groups. This comparison did not reveal chromosomal domains (Supplementary Fig. 8), supporting the hypothesis that the natural gene expression variation masks the effects observed with the discordant monozygotic twins. To validate this hypothesis further, we tested whether the gene expression fold change between trisomic and normal unrelated individuals was significantly higher in the upregulated GEDDs than the downregulated GEDDs when selecting the genes with the lowest gene expression variation. We observed a plateau of significant P values (centred around 1×10^{-30} and coefficient of variation <0.45) with approximately 2,500 genes in the upregulated and 3,500 genes in the downregulated GEDDs (Supplementary Fig. 6c, d). These results support the hypothesis that the natural gene expression variation occurring between unrelated individuals is extensive and that only stably expressed genes can unmask the effects observed with the discordant monozygotic twins.

GEDDs and previously described chromosomal domains

The domain organization of the mammalian chromosomes has been previously reported with the identification of the LADs[27]. Therefore we evaluated the correlation between LADs and GEDDs comparing the distribution of gene expression fold changes within and outside the LADs. We observed that the genes located within LADs were, on average, overexpressed in T1DS as opposed to the genes located outside of LADs (in inter-LADs (iLADs), $P = 8.9 \times 10^{-16}$) (Fig. 5a). A similar significant shift was observed in the Ts65Dn fibroblasts when we compared the gene expression fold changes inside and outside the mouse LADs[28] ($P = 4.6 \times 10^{-7}$) (Fig. 5b). The healthy monozygotic twins did not show any expression differences between LADs and iLADs ($P = 0.54$) (Fig. 5c). The LADs are mostly associated with an overall inhibition of gene expression[27,29,30]. In our experiment, the increased expression fold change inside LADs suggests a de-repression of the genes in the LADs of trisomic cells. Therefore, we reasoned that the interaction of the genome with the nuclear lamina might be disturbed in trisomic nuclei. We used the DamID (DNA adenine methyltransferase identification) method[31] to define the position of the LADs in the discordant twins' fibroblasts. The lamin B1 interaction scores along the chromosomes revealed highly correlated profiles of DNA-lamina interactions between T1DS and T2N (Spearman $\rho = 0.94$) (Fig. 5d, e and Supplementary Fig. 6d). We concluded that the overall topology of LADs is not detectably perturbed by the presence of an extra HSA21.

Furthermore, the LADs correlate with replication domains[32]. Early replicating domains contain mostly active genes and tend to localize centrally in the nucleus as opposed to late replicating domains that mainly contain less active genes at the nuclear periphery[33,34]. We assessed

©2014 Macmillan Publishers Limited. All rights reserved

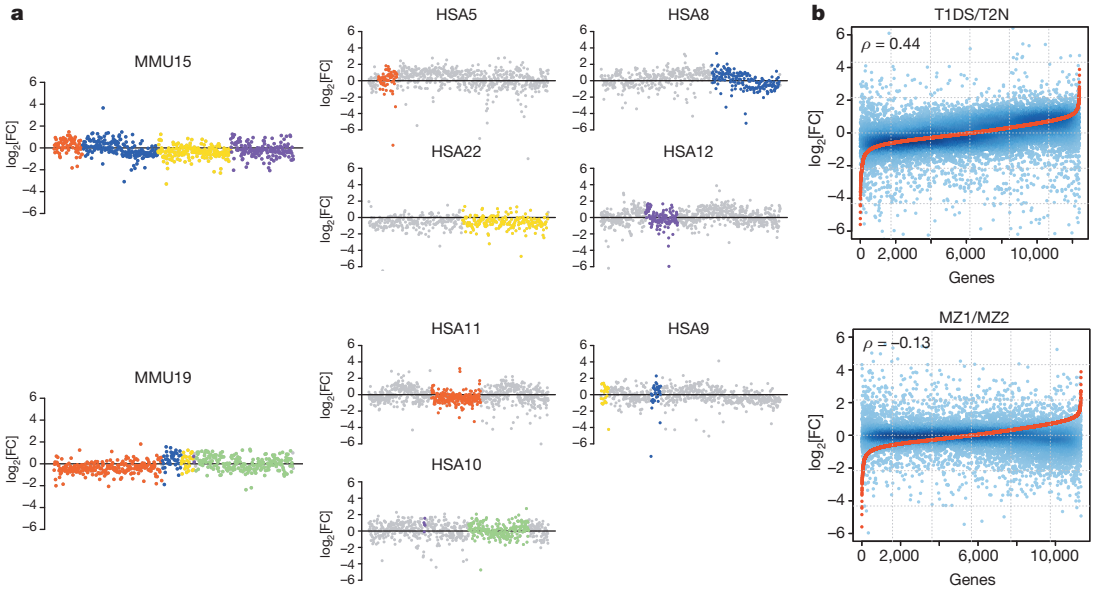


Figure 4 | GEDDs are conserved in Ts65Dn mouse model for Down's syndrome. **a**, Log_2 fold change of gene expression in the Ts65Dn fibroblasts (Ts65Dn/wild type) along mouse chromosomes 15 (MMU15, top panel) and 19 (MMU19, bottom panel). Genes are sorted according to their position on the chromosome (at equidistance). Colours represent the mouse/human syntenic blocks. The corresponding blocks on human chromosomes (T1DS/T2N Rep0) are shown with the same colours. **b**, Spearman correlation plot of log_2 expression fold changes between human (T1DS/T2N Rep0 in top panel and MZ1/MZ2 in bottom panel) and mouse (Ts65Dn/wild type) orthologous genes. Mouse genes (in red) are ranked in the ascending order of log_2[FC]. Human genes (in blue) are plotted in the same order. ρ = Spearman correlation.

the correlation between the replication time domains of human fetal fibroblasts[35] and the GEDDs described here. Most chromosomes show a strong correlation between both profiles, indicating that gene expression is increased in late replication domains and decreased in early replication domains of the T1DS cells (overall correlation $\rho = 0.6$, Fig. 6a and Supplementary Fig. 9).

Modification of the chromatin environment in GEDDs

Despite the fact that GEDDs are correlated to structural chromosomal domains, DamID experiments showed that the genome topology is not changed in T1DS fibroblasts. Thus, we reasoned that the alterations of gene expression might be related to epigenetic chromatin modifications within the chromosomal domains of trisomic cells.

We explored the DNA methylation changes in T1DS compared to T2N. We performed reduced representation bisulphite sequencing (RRBS)[36] on DNA from the twins' fibroblasts and evaluated the methylation percentage for each CpG dinucleotide in the genome. We then compared the methylation level of each gene between T1DS and T2N by considering the CpGs either in the gene body or in the promoter region. The profiles of gene methylation differences along the chromosomes were compared to the GEDDs. For some chromosomes, we observed a weak correlation between both patterns (Supplementary Fig. 10a, b). However, the overall degree of concordance between methylation and gene expression (overall $\rho = 0.29$) is not sufficient to correlate methylation status to GEDDs. We concluded that the alteration of gene expression cannot be fully explained by domain-wide changes of cytosine methylation in the trisomic fibroblasts.

The cellular transcription activity can also be influenced by changes of chromatin marks such as H3K4me3 that is known to correlate positively with gene expression[37]. Thus, we used chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) to compare the level of H3K4me3 in the discordant twins. We identified the genomic regions enriched for H3K4me3 in T1DS and T2N, and investigated the differences of signal intensity between all peaks common to both twins. The median of the resulting H3K4me3 log_2 fold change (log_2[FC]) for each gene was estimated to establish chromosomal profiles comparable to GEDDs. The comparison revealed a marked correlation between both patterns (overall $\rho = 0.93$) for all chromosomes except HSA13 and HSA18 (Fig. 6b and Supplementary Fig. 11). Those chromosomes do not show a domain pattern due to the overexpression of almost all of their genes. We conclude that the GEDDs observed in the twins correlate with modification of H3K4me3 in T1DS.

This observation raised the question of the general chromatin accessibility in the presence of an extra copy of HSA21. The DNase I hypersensitivity signal has been widely used to identify the accessible sites in the genome and to recognize the regulatory sequences[38]. We explored the chromatin accessibility in the twins' fibroblasts by performing DNase I hypersensitivity mapping. Interestingly, we detected an overall increased

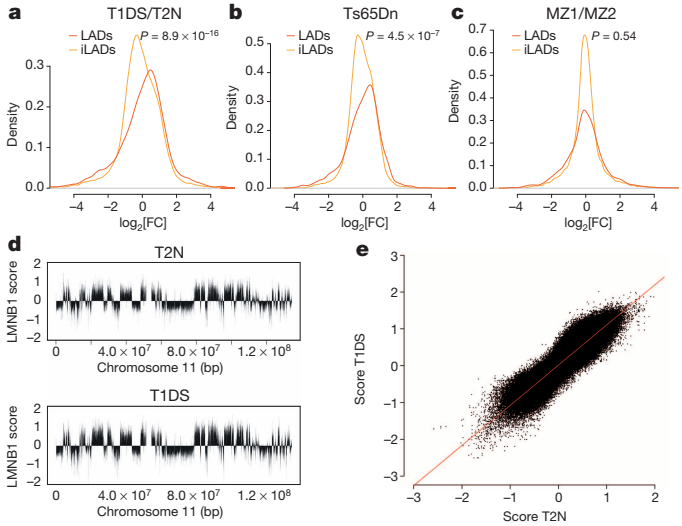


Figure 5 | Correlation with LADs. Distribution of log_2 expression fold changes for genes overlapping a LAD (in red) and genes located outside of a LAD (in yellow, inter-LADs (iLADs)). **a–c**, Profiles are shown for the discordant twins (Rep0) (**a**), the Ts65Dn mice (**b**) and the healthy monozygotic twins (**c**). P = Wilcoxon test P value. **d**, DamID map of lamin B1 (LMNB1) interaction scores for human chromosome 11 in T1DS (top panel) and T2N (bottom panel). **e**, Correlation plot of lamin B1 interaction scores for the entire genome between T1DS and T2N ($\rho = 0.94$). (Samples smoothed with a running median window of 11 probes.)

©2014 Macmillan Publishers Limited. All rights reserved

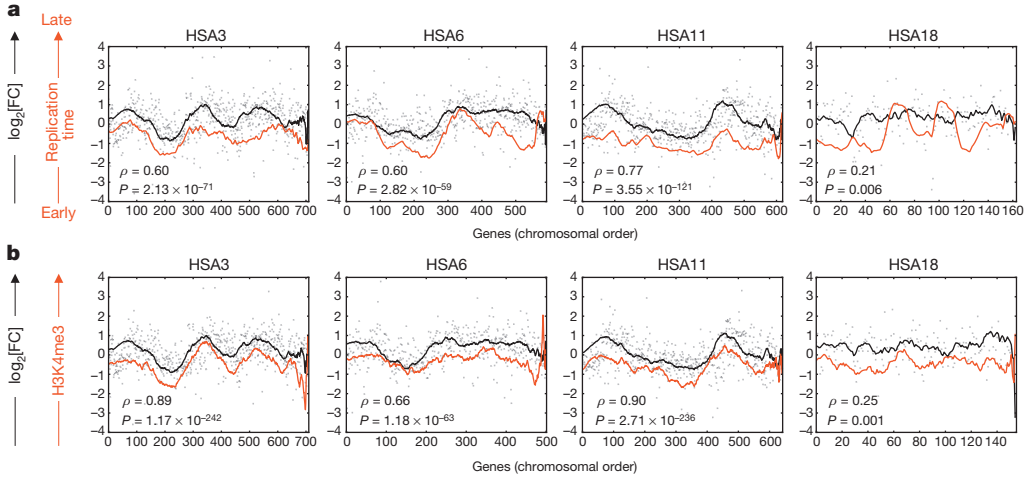


Figure 6 | Correlation with replication time domains and H3K4me3 mark. **a**, **b**, Comparison of the $\log_2$ expression fold change between T1DS and T2N (Rep0, in black) and the replication time (**a**) or the H3K4me3 difference (**b**) profiles (in red) along human chromosomes 3, 6, 11 and 18. ρ = Spearman correlation; P = associated P value.

genomic accessibility in the trisomic twin, particularly marked in the gene bodies, transcription start sites and promoter regions (Supplementary Fig. 6e). Because this modification may influence overall gene expression, we investigated the link between the changes of DNase I hypersensitivity profiles in the twins' fibroblasts and the GEDD pattern. Surprisingly, this comparison revealed that the two patterns were weakly linked and even tend to be inversely correlated (overall $\rho = -0.56$, Supplementary Fig. 12). Given the known correlation between DNase I hypersensitivity signal and marks for open chromatin, especially H3K4me3 (ref. 39), this observation is unexpected and may indicate some compensatory mechanisms for the regulation of gene expression in an open chromatin context.

Discussion

The transcriptome analysis of monozygotic twins discordant for trisomy 21 highlights the existence of chromosomal domains of gene expression dysregulation between trisomic and normal fibroblasts. This observation includes not only the expression of protein-coding genes but also lncRNAs (Supplementary Fig. 6f). We attribute this discovery to the use of samples from monozygotic twins, which eliminates the influence of individual genetic backgrounds. The investigation of gene expression level in each discordant twin independently revealed that the GEDD pattern could be attributed to a reduced dynamics of gene expression in the trisomic fibroblasts. The conservation of this transcriptome pattern in iPS cells suggests that the mechanism responsible for the dysregulation domains is not specific to fibroblasts but is maintained after dedifferentiation. The observed conservation also indicates that trisomy 21 has a consistent influence on the transcriptome of iPS cells.

We have demonstrated that GEDDs are conserved in the Ts65Dn mouse model for Down's syndrome. Remarkably, these domains are present in a genomic chromosomal context that is different in human and mouse. The conservation in mouse also suggests that the domain organization may be the consequence of the overexpression of one or more HSA21 gene(s), the orthologues of which are found on the MMU16 syntenic region. Alternatively, the domains could be related to the physical presence of the extra DNA material in the nucleus.

We have also shown that GEDDs significantly correlate with the lamina-associated and replication domains. Our results indicate that the trisomic cells undergo de-repression of the genes located at the nuclear periphery and repression of the early replicating active genes. The topology of LADs is not disturbed in trisomic cells; however, GEDDs are related to epigenetic modifications of the chromatin environment as shown by the almost perfect correlation with differential enrichment of H3K4me3 observed between T1DS and T2N. These modifications may occur to compensate general changes of chromatin accessibility as suggested by the DNase I hypersensitivity experiment.

Several HSA21 genes emerge as potential candidates for the epigenetic modification of chromosomal domains. *HLCS*, which is triplicated in the Ts65Dn mouse model, catalyses the binding of biotin to histones, and participates in chromatin condensation and gene repression[40,41]. Interestingly, HLCS is associated to the nuclear lamina[42] and can physically interact with chromatin-modifying proteins such as DNMT1 or MeCP2 (ref. 41).

The HSA21 protein HMGN1 also influences the structure and function of the chromatin through histone modifications[43] and is considered to be an important modulator of gene expression[44].

Other HSA21 proteins that are directly or indirectly involved in epigenetic mechanisms such as DYRK1A, BRWD1 and RUNX1 may influence the epigenetic architecture of the nucleus[45–47]. Additional experiments may help to explore their contribution to the GEDD pattern.

Furthermore, the correlation of GEDDs with the replication time domains and a prolonged permissive chromatin state in trisomic cells (as shown by our DHS experiments) raises the hypothesis of an excess of background transcription in trisomic cells. Several studies have shown that the cell cycle is extended in trisomic nuclei[48,49] without affecting the replication time[35]. This perturbation of the cell cycle may therefore result in a prolongation of chromatin access time and in a general increase of stochastic transcription[50].

We propose that the overexpression of one or more HSA21 candidate gene(s) modifies the chromatin environment of the nuclear compartments in trisomic cells. These modifications would lead to a general perturbation of the transcriptome that could explain some of the Down's syndrome phenotypes. Alternatively, the GEDDs could be the result of the additional chromosomal material of trisomy 21. The prediction of this latter hypothesis is that other trisomies may have a similar effect on the dysregulation of gene expression.

METHODS SUMMARY

Details on the following methods (and references within) can be found in the full Methods: cell culture and RNA preparation, mRNA sequencing, mapping and normalization, definition of chromosomal domains, expression level analysis, human/mouse comparison, lamina-associated domain analysis and DamID experiments, reduced representation bisulphite sequence (RRBS) library preparation, methylation computational analysis, H3K4me3 ChIP-sequencing, DNase I hypersensitivity mapping, and correlation analysis between $\log_2$ fold change in fibroblasts versus iPS $\log_2$[FC], IMR90 replication time, DNA methylation, H3K4me3 and DNase I hypersensitivity profiles.

©2014 Macmillan Publishers Limited. All rights reserved

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 September 2013; accepted 4 March 2014.

1. Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A. & Deutsch, S. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nature Rev. Genet.* **5,** 725–738 (2004).
2. Korenberg, J. R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl Acad. Sci. USA* **91,** 4997–5001 (1994).
3. Pritchard, M. A. & Kola, I. The "gene dosage effect" hypothesis versus the "amplified developmental instability" hypothesis in Down syndrome. *J. Neural Transm. Suppl.* **57,** 293–303 (1999).
4. Lyle, R. *et al.* Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. *Eur. J. Hum. Genet.* **17,** 454–466 (2009).
5. Shapiro, B. L. Down syndrome–a disruption of homeostasis. *Am. J. Med. Genet.* **14,** 241–269 (1983).
6. Letourneau, A. & Antonarakis, S. E. Genomic determinants in the phenotypic variability of Down syndrome. *Prog. Brain Res.* **197,** 15–28 (2012).
7. Davisson, M. T. *et al.* Segmental trisomy as a mouse model for Down syndrome. *Prog. Clin. Biol. Res.* **384,** 117–133 (1993).
8. Reeves, R. H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genet.* **11,** 177–184 (1995).
9. Chrast, R. *et al.* Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome. *Hum. Mol. Genet.* **9,** 1853–1864 (2000).
10. Ahn, K. J. *et al.* DYRK1A BAC transgenic mice show altered synaptic plasticity with learning and memory defects. *Neurobiol. Dis.* **22,** 463–472 (2006).
11. Costa, V. *et al.* Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21. *PLoS ONE* **6,** e18493 (2011).
12. Esposito, G. *et al.* Genomic and functional profiling of human Down syndrome neural progenitors implicates S100B and aquaporin 4 in cell injury. *Hum. Mol. Genet.* **17,** 440–457 (2008).
13. Lockstone, H. E. *et al.* Gene expression profiling in the adult Down syndrome brain. *Genomics* **90,** 647–660 (2007).
14. Sommer, C. A., Pavarino-Bertelli, E. C., Goloni-Bertollo, E. M. & Henrique-Silva, F. Identification of dysregulated genes in lymphocytes from children with Down syndrome. *Genome* **51,** 19–29 (2008).
15. Prandini, P. *et al.* Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am. J. Hum. Genet.* **81,** 252–263 (2007).
16. Deutsch, S. *et al.* Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum. Mol. Genet.* **14,** 3741–3749 (2005).
17. Dahoun, S. *et al.* Monozygotic twins discordant for trisomy 21 and maternal 21q inheritance: a complex series of events. *Am. J. Med. Genet. A.* **146A,** 2086–2093 (2008).
18. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature Methods* **9,** 1185–1188 (2012).
19. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
20. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
21. Ram, G. & Chinen, J. Infections and immunodeficiency in Down syndrome. *Clin. Exp. Immunol.* **164,** 9–16 (2011).
22. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291,** 1289–1292 (2001).
23. Gierman, H. J. *et al.* Domain-wide regulation of gene expression in the human genome. *Genome Res.* **17,** 1286–1295 (2007).
24. Hibaoui, Y. *et al.* Modelling and rescuing neurodevelopmental defect of Down syndrome using induced pluripotent stem cells from monozygotic twins discordant for trisomy 21. *EMBO Mol. Med.* **6,** 259–277 (2014).
25. Duchon, A. *et al.* Identification of the translocation breakpoints in the Ts65Dn and Ts1Cje mouse lines: relevance for modeling Down syndrome. *Mamm. Genome* **22,** 674–684 (2011).
26. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–511 (2013).
27. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951 (2008).
28. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
29. Zullo, J. M. *et al.* DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149,** 1474–1487 (2012).
30. Akhtar, A. & Gasser, S. M. The nuclear envelope and transcriptional control. *Nature Rev. Genet.* **8,** 507–517 (2007).
31. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nature Protocols* **2,** 1467–1478 (2007).
32. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107,** 139–144 (2010).
33. Gilbert, D. M. Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* **14,** 377–383 (2002).
34. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6,** e245 (2008).
35. Pope, B. D. *et al.* Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Hum. Mol. Genet.* **21,** 4162–4170 (2012).
36. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols* **6,** 468–481 (2011).
37. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
38. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
39. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
40. Singh, M. P., Wijeratne, S. S. & Zempleni, J. Biotinylation of lysine 16 in histone H4 contributes toward nucleosome condensation. *Arch. Biochem. Biophys.* **529,** 105–111 (2013).
41. Pestinger, V., Wijeratne, S. S., Rodriguez-Melendez, R. & Zempleni, J. Novel histone biotinylation marks are enriched in repeat regions and participate in repression of transcriptionally competent genes. *J. Nutr. Biochem.* **22,** 328–333 (2011).
42. Narang, M. A., Dumas, R., Ayer, L. M. & Gravel, R. A. Reduced histone biotinylation in multiple carboxylase deficiency patients: a nuclear role for holocarboxylase synthetase. *Hum. Mol. Genet.* **13,** 15–23 (2004).
43. Postnikov, Y. & Bustin, M. Regulation of chromatin structure and function by HMGN proteins. *Biochim. Biophys. Acta* **1799,** 62–68 (2010).
44. Zhu, N. & Hansen, U. Transcriptional regulation by HMGN proteins. *Biochim. Biophys. Acta* **1799,** 74–79 (2010).
45. Canzonetta, C. *et al.* DYRK1A-dosage imbalance perturbs NRSF/REST levels, deregulating pluripotency and embryonic stem cell fate in Down syndrome. *Am. J. Hum. Genet.* **83,** 388–400 (2008).
46. Huang, H., Rambaldi, I., Daniels, E. & Featherstone, M. Expression of the Wdr9 gene and protein products during mouse development. *Dev. Dyn.* **227,** 608–614 (2003).
47. Bakshi, R. *et al.* The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. *J. Cell. Physiol.* **225,** 569–576 (2010).
48. Contestabile, A. *et al.* Cell cycle alteration and decreased cell proliferation in the hippocampal dentate gyrus and in the neocortical germinal matrix of fetuses with Down syndrome and in Ts65Dn mice. *Hippocampus* **17,** 665–678 (2007).
49. Williams, B. R. *et al.* Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322,** 703–709 (2008).
50. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Rev. Genet.* **15,** 69–81 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Swiss National Science Foundation (SNF-144082), the European Research Council (ERC-249968), AnEUploidy and BluePrint EU grants, the Lejeune, and ChildCare foundations for supporting the S.E.A. laboratory. The laboratories of R.G. were supported by Spanish MICINN (BIO2011-26205) and ERC-294653, B.v.S. by NWO-ALW-VICI, Y.He. by CNRS, INSERM, University of Strasbourg and ANR-10-INBS-07, J.A.S. by NIH U54HG007010, and A.F. by Genico and Ernest Boninchi foundation. We thank S. Dahoun and J. L. Blouin for the discordant twins sample collection.

Author Contributions The project was coordinated by S.E.A. A.L. coordinated/undertook the main laboratory work. F.A.S. coordinated/undertook the main bioinformatics/statistical analyses. X.B. performed ChIP-seq experiments. M.R.S. performed DNA methylation, A.L., F.A.S., M.G., R.G. and D.G. processed NGS data. J.K. and B.v.S. performed DamID experiments. C.C. and Y.He. maintained the mouse colony and contributed mouse samples. R.T., R.S.S. and J.A.S. performed DNase experiments; Y.Hi. and A.F. derived the iPS cells; and K.P., D.R. R.G. and E.M. performed additional statistical analyses. E.F., M.G., C.G., A.V., M.G., L.F., C.B. and S.D. assisted with wet lab experiments and contributed to performing NGS experiments. The main findings were interpreted by S.E.A., A.L. and F.A.S., who also wrote the manuscript. All authors made comments on the manuscript.

Author Information All sequencing data have been deposited in the Gene Expression Omnibus (GEO) data repository under accession number GSE55426. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.A. (Stylianos.Antonarakis@unige.ch).

©2014 Macmillan Publishers Limited. All rights reserved

METHODS

Cell culture and RNA preparation. Forearm primary fetal skin fibroblasts were collected post mortem from the T1DS and T2N discordant twins at 16 fetal weeks, after IRB approval from the Ethical Committee of University Hospitals of Geneva, and written informed consent by both parents. Three replicate experiments (Rep0, Rep1 and Rep2) were performed from independent cultures of the same fibroblasts. The fourth replicate (Rep3) was performed using the RNA material of Rep0. Primary fibroblasts from MZ1 and MZ2 healthy monozygotic twins were derived from umbilical cord tissue collected as part of the GenCord project in the University Hospitals of Geneva[51]. Primary skin fibroblasts from trisomy 21 and normal unrelated individuals were taken from ref. 15 (GM08447, GM05756, GM00969, GM00408, GM02036, GM03377, GM03440, PM9726F, GM04616, AG07409, AG06872, AG06922, GM02767, AG08941, AG08942, D-S99124M). All primary fibroblasts were grown in DMEM GlutaMAX (Life Technologies 31966) supplemented with 10% fetal bovine serum (Life Technologies 10270) and 1% penicillin/streptomycin/fungizone mix (Amimed, BioConcept 4-02F00-H) at 37 °C in a 5% CO_2 atmosphere.

Mouse embryonic fibroblasts (MEFs) were obtained from 14.5-day-old mouse embryos. Embryos were collected, eviscerated, decapitated and then further treated with trypsin to obtain the MEFs. MEFs were cultured at least two passages in classical DMEM (Gibco) medium supplemented with antibiotics, glutamine and 10% fetal calf serum. Subconfluent MEFs were harvested in TRIzol lysis buffer (Life Technologies 15596).

Primary fetal skin fibroblasts isolated from T1DS and T2N (10 passages) were used to establish normal (Twin-N-iPSCs) and Down's syndrome (Twin-DS-iPSCs) induced pluripotent stem cells (iPS cells). These iPS cells were generated by transducing the parental fibroblasts with polycistronic lentiviral vectors expressing *OCT4*, *SOX2, KLF4* and *MYC* reprogramming factors[52,53]. The generated iPS cells were cultured on irradiated human foreskin fibroblasts (ATCC, CCD 1112Sk) that were mitotically inactivated by irradiation at 35 Gy. iPS cell colonies were maintained with daily changes in Knockout Dulbecco's Minimal Essential Medium (KO DMEM) supplemented with 20% KO serum replacement, 1 mM L-glutamine, 100 μM non-essential amino acids, 100 μM 2-mercaptoethanol, 50 U ml^{-1} penicillin and 50 mg ml^{-1} streptomycin (all from Gibco, Invitrogen) and 100 ng ml^{-1} human basic fibroblast growth factor (bFGF, Peprotech). These iPS cells were passaged by manual dissection of cell clusters in the presence of 10 μM ROCK-inhibitor Y-27632 (Sigma-Aldrich).

Total RNA from all cell types was collected using the TRIzol reagent (Life Technologies 15596) following the manufacturer's instructions. RNA isolation was performed after 10 or 13 passages of the twins' primary fibroblasts. iPS cells were grown for 17 passages before the RNA isolation step. RNA quality was verified on the Agilent 2100 Bioanalyzer (RNA 6000 Nano kit, 5067) and quantity was measured on a Qubit instrument (Life Technologies).

mRNA sequencing, mapping and normalization. mRNA-seq libraries were prepared from 4 to 10 μg of total RNA using the Illumina mRNA-seq Sample Preparation kit (RS-100-0801), following the manufacturer's instructions. Libraries were sequenced on the Illumina HiSeq 2000 (Rep0 and Rep1) and on the Illumina Hiseq 2500 (Rep2 and Rep3) using paired-end sequencing 2 × 100 bp. Reads were mapped against the human (hg19) or mouse (mm9) genomes using the default parameters of the GEM mapper through the GRAPE pipeline[18,54]. An independent mapping was performed with TopHat using default parameters. Quantile normalization was performed on the RPKM data (reads per kilobase per million) for all the human samples on one side and for the mouse samples on the other side. Differential expression analysis was performed on the raw read counts using the default parameters of EdgeR on the genes expressed in at least two samples[19].

Definition of chromosomal domains. For each gene, we determined the $\log_2$ expression fold change ($\log_2$[FC]) by calculating the ratio of normalized expression values (after quantile normalization) between the trisomic and the euploid samples (that is, a positive $\log_2$[FC] reflects the overexpression of the gene in the trisomic sample). For the healthy monozygous twins the $\log_2$[FC] was based on the ratio between MZ1 and MZ2. For the unrelated individuals, the $\log_2$[FC] was based on the ratio between the mean value of all trisomy 21 individuals and the mean value of all normal individuals. For all the comparisons, only the genes with at least 0.1 RPKM in the normal individual(s) (T2N, average value of MZ1 and MZ2 or average value of the normal unrelated individuals) were considered. The $\log_2$[FC] values were plotted equidistantly based on the gene positions along the chromosomes. We used the lowess function in R (http://www.r-project.org) to smooth the $\log_2$[FC] data (smoother span 3 to 30%) and identify upregulated and downregulated domains. An upregulated domain was defined as a set of at least two consecutive genes with positive smoothed $\log_2$[FC] values. On the contrary, a downregulated domain was defined as a set of at least two consecutive genes with negative smoothed $\log_2$[FC] values.

Expression level analysis. The genes were classified according to their expression level (RPKM value after quantile normalization) in three categories: low ($0.1 < RPKM < 10$), medium ($10 < RPKM < 100$) or high ($RPKM > 100$) expression. In the discordant twins comparison, we used the expression level of the normal twin (T2N) as a reference. In the healthy monozygous twins comparison, we used the average value between MZ1 and MZ2 to assess the gene expression level. For each category, the distribution of $\log_2$[FC] was plotted (density function in R). The difference of distribution between the discordant twins and the healthy monozygous twins was assessed by the Wilcoxon test.

Gene expression dynamics for each sample has been evaluated by calculating the standard deviation of the lowess smoothed $\log_2$ expression of each gene sorted according to chromosomal coordinates.

Unrelated individuals analysis. We calculated the coefficient of variation (CV) of each gene over all 8 + 8 samples as the standard deviation of its relative expression divided by the mean. The genes with CV above a decreasing threshold were iteratively discarded. At each step, we evaluated the expression fold change of all the remaining genes by considering all possible pairwise comparisons of the 8 trisomic and 8 control samples. We applied a non-parametric Mann–Whitney *U*-test between the fold change values obtained for the genes located in the upregulated GEDDs and the genes located in the downregulated GEDDs.

Human/mouse comparison. The mouse/human syntenic block coordinates were downloaded from the "Comparative Genomics" track of the UCSC genome browser (http://genome.ucsc.edu). Only the largest mouse syntenic blocks (>1 Mb) were conserved for the comparative analysis. $\log_2$[FC] of gene expression between Ts65Dn and wild-type mice were plotted along the mouse chromosomes. Each gene was assigned to a syntenic block (total overlap between the gene coordinates and the block coordinates) and coloured accordingly. The corresponding blocks in human were shown with the same colours in the GEDD plots.

The list of mouse and human orthologous genes was obtained from the Mouse Genome Informatics Database project, The Jackson Laboratory, Bar Harbour, Maine (ftp://ftp.informatics.jax.org/pub/reports/HMD_HumanPhenotype.rpt). We compared the $\log_2$[FC] of expression obtained in the mouse with the values obtained for their human orthologues. The comparison was done first with the discordant twins and then with the healthy monozygous twins. The degree of similitude between the 2 sets was given by the Spearman correlation.

LADs analysis and DamID. The coordinates of human LADs were taken from ref. 27, and the mouse LAD coordinates were taken from ref. 28 (mouse embryonic fibroblasts). A gene was assigned to a LAD if its coordinates overlap from at least 1 bp with the LADs coordinates. The distribution of $\log_2$[FC] for the overlapping genes was compared to the distribution of $\log_2$[FC] for the non-overlapping genes using a Wilcoxon test. DamID was performed as previously described[31].

Reduced representation bisulphite sequencing (RRBS) library preparation. Reduced representation bisulphite sequencing (RRBS) libraries were made according to ref. 36 with some modifications. In short, the Qiagen DNeasy Blood and Tissue kit was used to extract DNA from the twin fibroblast cells. We then digested 2 μg genomic DNA with 2 μl 20 U μl^{-1} MspI restriction enzyme that cuts DNA regardless of cytosine methylation status at CCGG sequence in 5 μl NEB buffer 4 in a total reaction volume of 50 μl. This reaction was incubated for 18 h at 37 °C followed by heat inactivation at 80 °C for 20 min to generate DNA fragments containing CpG dinucleotides at the end. Phenol extraction and ethanol precipitation steps were used to clean up DNA from enzymatic reaction. DNA fragments were subsequently end-repaired and A-tailed by adding 2 μl dNTP mix (10 mM), 2 μl 5 U μl^{-1} Klenow fragment, and 4 μl of NEB2 in a total reaction volume of 40 μl. This reaction was incubated at 30 °C for 20 min, followed by 20 °C for 37 min. Heat inactivation was done at 75 °C for 20 min. The end-repaired and A-tailed DNA was purified by phenol extraction and ethanol precipitation steps and ligated to the methylated version of Illumina adapters: ilAdap Methyl PE1 (ACACTCTTTCCC TACACGACGCTCTTCCGATC*T) and ilAdap Methyl PE2 (GATCGGAAGAG CGGTTCAGCAGGAATGCCGA*G); all Cs are methylated, 5′ phosphate, asterisk indicates phosphorothioate bond. The ligation was mediated by adding 1 μl T4 DNA ligase (2,000 U μl^{-1}), 2 μl T4 ligase buffer (10×), and 1 μl methylated adapters (paired-end adapters) (15 μM) in a total reaction volume of 20 μl. The reaction was incubated at 16 °C overnight. The adaptor-ligated DNA was purified by phenol extraction and ethanol precipitation steps and dissolved in 15 μl EB buffer. Size selection of the adaptor-ligated DNA fragments (170 bp to 350 bp) was done by electrophoresing the 15 μl ligation reaction in a 2.5% NuSieve GTG agarose gel (Lonza). We subsequently purified the DNA using a Qiagen Qiaquick Gel Extraction kit as described in the manufacturer's instructions, except that we eluted the purified DNA from the column using 25 μl buffer EB. We then used 20 ml of this purified DNA in the sodium bisulphite conversion, which was performed using the QIAGEN Epitect Bisulphite kit (Qiagen) per manufacturer's recommended protocol with the following modifications: incubation after the addition of bisulphate conversion reagents was conducted in a thermocycler with the following conditions:

©2014 Macmillan Publishers Limited. All rights reserved

95 °C for 5 min, 60 °C for 25 min, 95 °C for 5 min, 60 °C for 85 min, 95 °C for 5 min, 60 °C for 175 min (95 °C for 5 min and 60 °C for 90 min) × 6. The bisulphate-converted DNA was eluted two times in 20 µl of EB buffer yielding 40 µl at the end. Illumina PCR primers PE1 and PE2 were used for the final library amplification. The sequences of PCR primers are as follow: ilPCR PE1 (AATGATACGGCGACCACC GAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T) and ilPCR PE2 (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCC TGCTGAACC GCTCTTCCGATC*T), where an asterisk indicates a phosphorothioate bond. The purified bisulphate-treated DNA was then PCR-amplified in a reaction containing 100 µl KAPA2G Robust DNA Polymerase mixture (Kapabiosystems), 40 µl bisulphite-treated DNA, 0.5 mM ilPCR PE1, 0.5 mM ilPCR PE2 DNA oligonucleotide primers in a total reaction volume of 200 µl. The PCR mixtures were divided among eight PCR tubes and were incubated at 95 °C for 2 min, followed by 20 cycles of (95 °C for 20 s and 65 °C for 30 s and 72 °C for 30 s) and 72 °C for 7 min. The PCR products were pooled and purified by adding 360 µl AMPure magnetic beads (Agencourt Bioscience). We confirmed the amplification and correct product size range by running one-fifth of the reaction on a 2% agarose gel. We quantified the purified product using the Qubit fluorometer (Invitrogen). We also checked the template size distribution by running an aliquot of the library on an Agilent Bioanalyzer (Agilent 2100 Bioanalyzer). We then diluted each library to 10 nM and proceeded to sequence each library (pair-ended 100 bp) in a single lane on the Illumina HiSeq 2000 according the manufacturer's instructions. We also ran one lane of a standard (non-RRBS) library (Exome) in the same flow cell as the control lane and used this lane (rather than the RRBS lane) to calculate the dye matrix for base calling. Also, 5% phiX genomic DNA (control) was spiked into each lane. We also limited the cluster number for RRBS libraries to an 80% optimized cluster number per tile on Illumina HiSeq 2000. Image capture, analysis and base calling were performed using Illumina's CASAVA 1.7.

Methylation computational analyses. We obtained 121.7 million reads for T1DS and 104.3 million reads for T2N of which after trimming of reads for base quality and adaptor contamination (trim_galore wrapper http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) 59% and 59.6% of reads were uniquely mapped against human genome hg19 using Bismark aligner (-n 3 -l 20 -I 20)[55]. We used MethylKit R package[56] to calculate methylation percentages for each CpG. Per cent methylation values for CpG dinucleotides are calculated by dividing number of methylated Cs by total coverage on that base. CpGs with at least 20× read coverage and at least a Phred score threshold of 20 were retained for calling CpG methylation. 3,984,938 CpGs in T1DS and 3,690,212 CpGs in T2N met these criteria of which 2,598,760 CpGs are covered in both samples.

H3K4me3 ChIP sequencing. Primary human skin fibroblasts from both twins were grown in 10 cm dishes without reaching confluence. 21 million cells were crosslinked by adding enough formaldehyde for final concentration 0.5% to the growing media at room temperature. Dishes were incubated for 10 min and the reaction was then quenched by adding glycine to a final concentration of 125 mM. The crosslinked cells were collected and stored at −80 °C. For cell lysis and sonication, the pellets were re-suspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0) and sonicated in a Covaris S2 instrument (duty cycle 5%, intensity 5, 200 cycles per burst, 6 min).

Chromatin from 7.5 million cells was incubated with 10 µl of H3K4me3 antibody (Abcam ab8580 0.5 mg ml^{-1}) coupled to IgG magnetic beads (Cell Signaling) at room temperature for 1 h and subsequently at 4 °C for another hour. The magnetic beads were washed five times with LiCl wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and one time with TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA). Subsequently, the bound DNA was eluted at 65 °C in elution buffer (1% SDS, 0.1 M $NaHCO_3$) for 1 h and then separated from the beads with a magnetic stand. The immunoprecipitated DNA was incubated overnight at 65 °C with 2 µl Proteinase K to reverse crosslink. The samples were then purified with the QIAquick purification kit (Qiagen).

ChIP-seq libraries were prepared using a ChIP-seq DNA Sample Prep kit from Illumina with the following modifications to the protocol: mRNA adaptor indexes from the TruSeq RNA kit were used instead of the Adaptor oligo mix. The enrichment PCR was carried out with PCR reagents from the TruSeq mRNA kit from Illumina. The enriched library was purified with AMPure magnetic beads (Agencourt), its concentration verified with Qubit (Invitrogen) and the distribution and size of fragments with a Bioanalyzer (Agilent). Four samples were pooled in equimolar proportion and sequenced in a HiSeq2500 (single read, 36 bp). The obtained reads were demultiplexed with Illumina CASAVA 1.8 software and mapped to hg19 with BWA. Peaks were called using unique tags with HOMER (http://sdcsb.ucsd.edu/resources/homer/). Peak comparison between twins was performed with HOMER and in house scripts.

DNase I hypersensitivity mapping. DNase I hypersensitivity mapping and analysis were performed as previously described[38,57]. DNase I hypersensitivity density plots were generated by counting the reads covering each gene body ± 5 kb and normalizing by the respective gene length. Each gene body has been divided in 100 non-overlapping bins +50 bins for ± 5 kb extensions.

Correlation analysis between log_2[FC] fibroblast versus iPS log_2[FC], IMR90 replication time, DNA methylation, H3K4me3 and DNase I hypersensitivity profiles. iPS log_2[FC] chromosomal profiles have been obtained from RNA-seq data as previously described.

Replication time domains of IMR90 fetal fibroblasts were downloaded from the ReplicationDomain database (http://www.replicationdomain.org). Gene replication times have been estimated for each gene considering the median of each overlapping replication time domain and ordered according to chromosomal coordinates.

H3K4me3 profiles have been obtained as follows. For each gene body ±1,000 bp, the overlapping H3K4me3 ChIP-seq peaks were isolated and the log_2[FC] of the size of the matching peaks (as provided by HOMER) of the trisomic and the healthy monozygotic twin were calculated. The median of the resulting H3K4me3 log_2[FC] for each gene was eventually retained to compose the chromosomal profiles according to gene coordinates.

Similarly, DNase I hypersensitivity profiles have been obtained considering the log_2[FC] of the size of the overlapping DNase peaks for each gene body ±1,000 bp. The median of the resulting DNase log_2[FC] for each gene was retained to compose the chromosomal profiles according to gene coordinates.

DNA methylation profiles in gene bodies and promoters have been obtained from averaging per cent methylation values for CpG dinucleotides in gene bodies and in promoter regions encompassing the transcription start sites [TSS−2000 bp, TSS+100 bp], respectively.

Locally weighted scatterplot smoothing (lowess) with 30% bandwidth has been applied before calculating the Spearman correlation between fibroblast log_2[FC] gene expressions and iPS log_2[FC], gene replication times, DNA methylation, DNase I hypersensitivity and H3K4me3 profiles for each chromosome. Significance of each chromosomal correlation has been estimated by 10^6 random permutations. When the obtained empirical P value was zero, we reported the theoretical one.

The above-mentioned chromosomal lowess smoothed profiles have been concatenated from HSA1 to HSAX to calculate the overall Spearman correlation with gene expression. |Overall correlations| >0.2 are all significant ($P < 10^{-300}$).

51. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325,** 1246–1250 (2009).
52. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
53. Grad, I. *et al.* NANOG priming before full reprogramming may generate germ cell tumours. *Eur. Cells Mat.* **22,** 258–274 (2011).
54. Knowles, D. G., Roder, M., Merkel, A. & Guigo, R. Grape RNA-Seq analysis pipeline environment. *Bioinformatics* **29,** 614–621 (2013).
55. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
56. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13,** R87 (2012).
57. John, S. *et al.* Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* Ch. 27, Unit 21 27 (2013).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13023

Brainstem nucleus MdV mediates skilled forelimb motor tasks

Maria Soledad Esposito[1,2], Paolo Capelli[1,2] & Silvia Arber[1,2]

Translating the behavioural output of the nervous system into movement involves interaction between brain and spinal cord. The brainstem provides an essential bridge between the two structures, but circuit-level organization and function of this intermediary system remain poorly understood. Here we use intersectional virus tracing and genetic strategies in mice to reveal a selective synaptic connectivity matrix between brainstem substructures and functionally distinct spinal motor neurons that regulate limb movement. The brainstem nucleus medullary reticular formation ventral part (MdV) stands out as specifically targeting subpopulations of forelimb-innervating motor neurons. Its glutamatergic premotor neurons receive synaptic input from key upper motor centres and are recruited during motor tasks. Selective neuronal ablation or silencing experiments reveal that MdV is critically important specifically for skilled motor behaviour, including accelerating rotarod and single-food-pellet reaching tasks. Our results indicate that distinct premotor brainstem nuclei access spinal subcircuits to mediate task-specific aspects of motor programs.

Initiation of natural movement depends on the function of descending pathways to the spinal cord. This critical need is strikingly obvious in patients with complete spinal cord injury who lack the ability to move muscles controlled by spinal segments below the lesion, despite the presence of functional circuits in the spinal cord. Descending motor control pathways are at the core of supporting different forms of movement spanning from repetitive basic locomotor tasks such as walking[1–4] to sophisticated fine motor tasks like object manipulation[5–9]. Classical studies provide evidence that electrical stimulation of diverse areas in the brainstem can elicit a variety of movement sequences mediated by the reticular formation[1,2,4,10–13]. However, neuronal subpopulations and descending circuit modules linked to these diverse functions remain largely undefined.

Descending motor pathways with projections to the spinal cord operate in a dual mode of synaptic communication. They can contact spinal interneurons to indirectly influence motor neuron activity, and they can establish direct synaptic connections with motor neurons representing the shortest line of communication to produce motor output[6,7,14–18]. Evolutionary studies on motor cortex connectivity to spinal motor neurons have provided evidence for greater direct cortical input to motor neurons in species with highly skilled motor behaviours[5–7], supporting the idea that evolutionary changes in motor neuron access may influence the action task repertoire an animal can perform. In addition, within any mammalian species, forelimbs are superior to hindlimbs in the execution of sophisticated tasks, raising the question of whether interlimb circuitry comparison of descending input from brainstem to motor neurons may reveal differences in connectivity matrices related to function.

Here we use a trans-synaptic virus approach to visualize and compare connectivity patterns of neurons in the brainstem to forelimb- or hindlimb-innervating motor neurons in the mouse spinal cord, revealing striking differences of descending input to distinct spinal motor neurons. The medullary reticular formation ventral part (MdV) is a key brainstem area specifically connecting to a subset of forelimb-innervating spinal motor neurons. We show that MdV neurons receive input from upstream motor control centres and are essential for efficient performance of skilled motor tasks. Our findings provide a circuit- and subpopulation-level connectivity map for descending pathways regulating limb motor control and put forward a model in which distinct brainstem hubs differentially address spinal circuits to control motor actions.

Brainstem neurons for limb motor control

To reveal the identity of descending brain areas with direct synaptic connections to motor neurons innervating forelimb or hindlimb muscles, we used monosynaptic rabies virus technology[19–21] for retrograde infection of motor neurons in the spinal cord[22,23] (Fig. 1a). Unilateral virus injections into forelimb or hindlimb muscles resulted in reliable labelling of brainstem neurons premotor to these motor neurons. To identify differences in the distribution of forelimb and hindlimb brainstem premotor neurons, we developed a digital three-dimensional brainstem model and assigned identity tags (Fig. 1b; Extended Data Fig. 1 and Supplementary Videos 1–3) on the basis of brainstem substructures in which neurons resided[24]. We found that the diversity of highlighted brainstem nuclei overall was higher for forelimb than hindlimb premotor neurons, but significant representation biases were detected in both directions (Fig. 1c–f). Hindlimb premotor neurons showed relative overrepresentation in vestibular nuclei and in the spinal vestibular nucleus (Fig. 1e, f and Extended Data Fig. 2). In contrast, forelimb premotor neurons showed a strong bias in two bilateral laterally located neuron clusters (spinal trigeminal nucleus and parvicellular reticular nucleus) as well as in a predominantly ipsilaterally residing medial neuron cluster in the caudal medulla, an area previously named medullary reticular formation ventral part (MdV) (Fig. 1c–f; Extended Data Figs 2 and 3). Other brainstem areas showed approximately equal occupancy by forelimb and hindlimb premotor neurons (Fig. 1e, f and Extended Data Fig. 3). Prominent examples were the magnocellular reticular nucleus and the gigantocellular reticular nucleus (Fig. 1e, f and Extended Data Fig. 3), in line with previous work suggesting a role for these substructures in locomotor control[2]. Neurons assigned to brainstem substructures could also be distinguished based on neuronal morphology and cell soma area measurements (Extended Data Fig. 4). Taken together, these findings support the notion that diverse brainstem premotor populations differentially regulate forelimb and hindlimb

[1]Biozentrum, Department of Cell Biology, University of Basel, Basel 4056, Switzerland. [2]Friedrich Miescher Institute for Biomedical Research, Basel 4058, Switzerland.

©2014 Macmillan Publishers Limited. All rights reserved

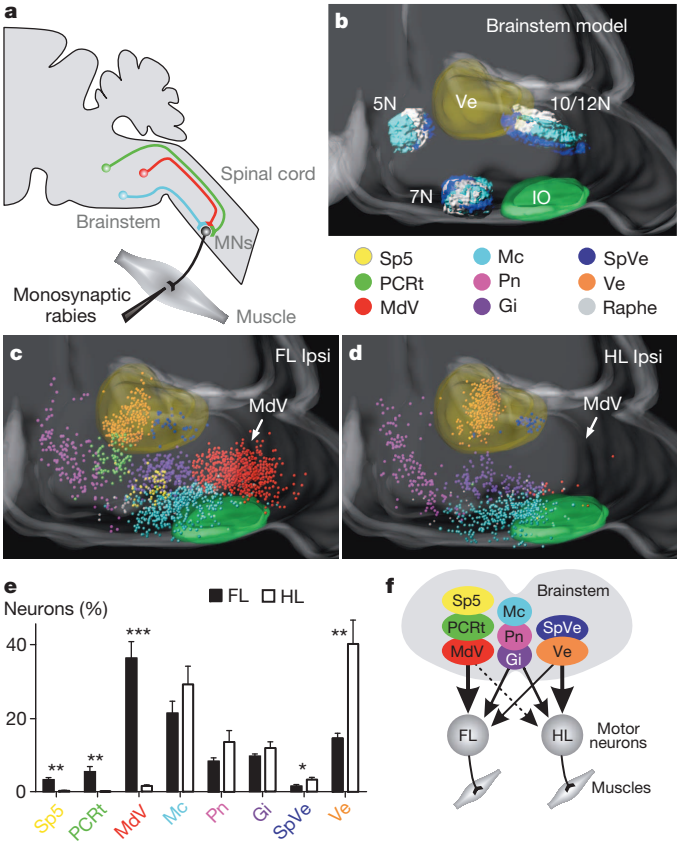


Figure 1 | Brainstem premotor distribution. **a**, Scheme illustrating method used to visualize premotor brainstem neurons labelled with trans-synaptic monosynaptic rabies viruses upon limb muscle injections. **b**, Brainstem model for three-dimensional reconstructions. Depicted are cranial motor nuclei 5N, 7N, 10/12N, vestibular nuclei (Ve) and inferior olive (IO). Gi, gigantocellular reticular nucleus; Mc, magnocellular reticular nucleus; MdV, medullary reticular nucleus ventral part; PCRt, parvicellular reticular nucleus; Pn, pontine reticular nucleus; Sp5, spinal trigeminal nucleus; SpVe, spinal vestibular nucleus; Ve, vestibular nucleus. **c**–**f**, Sagittal ipsilateral (Ipsi) views of three-dimensional brainstem reconstructions for forelimb (FL, **c**) and hindlimb (HL, **d**) premotor neurons ($n = 5$ mice each), quantification (**e**) and summary diagram (**f**) of ipsilateral connection specificity.

motor neurons, with some showing strongly biased access to one or the other class of motor neurons (Fig. 1f). To begin to understand the role of populations with differential motor neuron access, we chose one of the most biased examples for further investigation—the forelimb-dominated premotor population in the MdV.

MdV connections to forelimb motor neurons

To determine whether all MdV neurons establish synaptic connections with cervical spinal motor neurons, we next analysed neuronal diversity within MdV using genetic approaches to separately label glutamatergic (vGlut2^{ON}) or GABA (γ-aminobutyric acid)/glycinergic (vGAT^{ON}) neurons[25,26]. We found that glutamatergic and inhibitory neurons are almost equally represented and intermingled (Extended Data Fig. 5; vGlut2^{ON}, 45 ± 3%; vGAT^{ON}, 63 ± 3%; less than 1% cholinergic neurons). However, forelimb-premotor MdV neurons were almost exclusively glutamatergic (vGlut2^{ON}, 93 ± 1%; vGAT^{ON}, 6 ± 2%), demonstrating differential connectivity patterns of MdV neurons with distinct neurotransmitter phenotype (Extended Data Fig. 5).

This finding allowed us to specifically mark forelimb-connecting MdV neurons by stereotaxic injections of loxP-flanked conditional adeno-associated viruses (AAV-flex, flip excision switch) into *vGlut2^{Cre}* mice. To determine connection specificity between MdV and different forelimb-innervating motor neuron pools, we tagged synapses from vGlut2^{ON} MdV neurons by selective expression of a fusion protein between synaptophysin and enhanced green fluorescent protein (AAV-flex–Syn–GFP), combined with retrograde labelling of motor neurons from corresponding muscles (Fig. 2a, b). MdV input to motor neurons exhibited a highly discriminatory pattern to different motor neurons. We found high synaptic density contacts to biceps, extensor carpi radialis and extensor digiti quarti motor neurons, but only sparse input to most other forelimb motor pools analysed (for example, triceps, palmaris longus and flexor carpi radialis) (Fig. 2c–e)[27,28] or hindlimb-innervating motor neurons (Fig. 2c).

The prominent synaptic input difference from MdV to biceps and triceps motor neurons prompted us to also assess the connectivity profile of other brainstem premotor neurons to these functionally antagonistic motor neuron pools by monosynaptic rabies technology. The retrograde assay also showed that MdV neurons have a strong connection bias towards biceps compared to triceps motor neurons, whereas analysis of magnocellular reticular nucleus and gigantocellular reticular nucleus neurons showed equal connectivity ratios to biceps or triceps motor neurons (Fig. 2f–h and Extended Data Fig. 6). In contrast, neurons residing in the parvicellular reticular nucleus, spinal trigeminal nucleus, pontine reticular nucleus and vestibular nucleus showed connectivity profiles biased towards triceps motor neurons (Fig. 2f–h and Extended Data Fig. 6), in support of previous work demonstrating preferential connections of vestibular nucleus neurons to extensor motor neurons[16]. Together, these findings demonstrate that glutamatergic MdV neurons show high synaptic specificity directed towards only a few forelimb motor neuron pools, and that other brainstem nuclei show synaptic specificity patterns opposite to MdV (Fig. 2h), thus unravelling a diverse but specific connectivity matrix for descending brainstem input to functionally distinct motor neuron pools.

MdV connects to spinal interneurons

The anterograde tracing approach from vGlut2^{ON} MdV neurons allowed us to visualize axonal trajectory and synaptic pattern in the spinal cord, by using a combination of cytosolic and synaptic fluorescent tracers (AAV-flex–Tomato and AAV-flex–Syn–GFP). We found that vGlut2^{ON} MdV neurons have a descending trajectory mostly confined to the ipsilateral dorsolateral funiculus terminating in an ipsilateral central domain (Extended Data Fig. 7). To probe for the neurotransmitter identity of MdV-connected spinal interneurons, we initiated monosynaptic rabies virus spread from vGlut2^{ON} or vGAT^{ON} spinal interneurons by applying an established method to visualize presynaptic neurons from genetically marked neurons[29]. We performed intraspinal injections targeting the caudal cervical (C6–C8) spinal cord in *vGlut2^{Cre}* or *vGAT^{Cre}* mice, sequentially injecting AAV-flex–TVA/G (avian sarcoma and leukosis virus subtype A receptor and rabies glycoprotein) and rabies coated by the avian sarcoma leukosis virus envelope protein (EnvA).

We detected MdV neurons with monosynaptic connections to both of these populations (Extended Data Fig. 7). These findings show that MdV neurons not only contact specific motor neuron pools at forelimb levels, but also connect to excitatory and inhibitory interneurons, establishing both direct and indirect pathways to spinal motor neurons.

MdV controlled by upstream motor centres

To understand the role of MdV neurons in behaviour, we next determined the identity of synaptic input regulating forelimb-premotor neurons in the MdV. Because cortical input is implicated in the control of skilled motor behaviour[6,7], we first assessed whether premotor MdV neurons receive synaptic input from forebrain neurons (Extended Data Fig. 8a–c). In mice with genetically tagged forebrain synapses[26,30], forelimb premotor MdV neurons receive synaptic input at more than threefold higher density (Syn–GFP^{ON}/vGlut1^{ON}) than gigantocellular reticular nucleus or pontine reticular nucleus premotor neurons (Extended Data Fig. 8c), known to receive cortical input[18].

To determine in a more unbiased way the identity of presynaptic populations regulating vGlut2^{ON} MdV neurons, we applied a retrograde

©2014 Macmillan Publishers Limited. All rights reserved

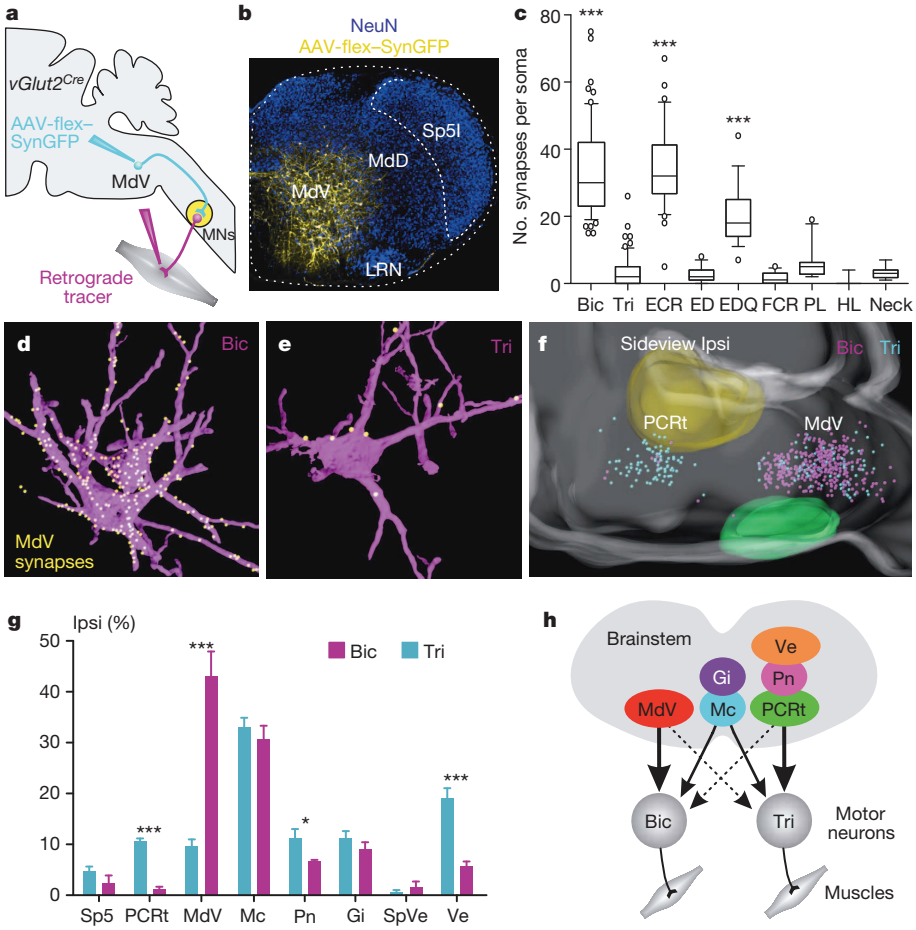


Figure 2 | Selective motor pools contacted by MdV. **a**, Scheme of experimental setup for MdV injection of AAV-flex–Syn–GFP virus and retrograde motor neuron marking. **b**, Injection site in MdV of *vGlut2*^{Cre} mice (MdD, medullary reticular nucleus, dorsal part; LRN, lateral reticular nucleus). **c**, Synaptic input from vGlut2^{ON} MdV neurons to selected motor neuron pools. Bic, biceps; Tri, triceps; ECR, extensor carpi radialis; ED, extensor digitorum communis; EDQ, extensor digiti quarti; FCR, flexor carpi radialis; PL, palmaris longus, flexor. For muscle assignment, see refs 27, 28. **d**, **e**, Biceps and triceps motor neuron reconstructions with input from vGlut2^{ON} MdV neurons (yellow, opposed terminals). **f**–**h**, Analysis of biceps and triceps premotor PCRt and MdV populations in sagittal ipsilateral view of brainstem reconstructions (**f**), quantification of ipsilateral premotor neurons ($n = 5$ mice each) (**g**) and summary diagram of connection specificity (**h**).

trans-synaptic labelling initiated from vGlut2^{ON} MdV neurons. We found that several prominent sites rostral to the hindbrain provide synaptic input to vGlut2^{ON} MdV neurons (Fig. 3a), including layer V pyramidal neurons in M1 motor cortex (Fig. 3a, b), deep layer vGlut2^{ON} neurons in the superior colliculus (Fig. 3a, c), vGlut2^{ON} neurons in subsets of deep cerebellar nuclei (Fig. 3a, d, e), vGlut2^{ON} neurons in the red nucleus (Fig. 3a, f), and inhibitory neurons in the paratrochlear nucleus (Extended Data Fig. 8d). Input was also detected from the reticular formation of the brainstem (Extended Data Fig. 8e–g). Together, these findings demonstrate that glutamatergic MdV neurons receive synaptic input from several important upstream motor control centres.

MdV recruited during motor tasks

To address whether MdV neurons are recruited during motor behaviour, we carried out two different motor tasks and measured the expression of the immediate early gene product c-Fos. As tasks we chose the basic locomotor task of treadmill locomotion and a skilled motor task on an accelerating rotarod, in which mice learn to stay on a rotating rod with accelerating speed[31–33]. We found that the number of neurons expressing c-Fos in MdV was significantly higher in both paradigms compared to untrained mice, with a preferential increase in vGlut2^{ON} MdV neurons (Extended Data Fig. 9). Together and consistent with previous results[34], MdV neurons are engaged in execution of motor programs as monitored by c-Fos expression.

MdV required for skilled motor tasks

To assess the role of MdV in the performance of motor tasks, we expressed human diphtheria toxin receptor (DTR)[35] specifically in vGlut2^{ON} MdV neurons by stereotaxic injections of a loxP-flanked conditional AAV (AAV-flex–DTR), targeting 82 ± 6% of vGlut2^{ON} MdV neurons around the injection site. Intraperitoneal diphtheria toxin injection resulted in complete ablation of these neurons 14 days after treatment (MdV-DTR mice). Comparing MdV-DTR and control groups, we did not observe defects in performing simple motor tasks including general cage behaviour, locomotion in an open field arena, running on a home cage wheel as well as even-spaced ladder walking (Fig. 4a–c; data not shown). We therefore next assessed performance in two different skilled motor tasks.

First, we measured the latency that MdV-DTR and control mice stayed on an accelerating rotarod[31,36] (Fig. 4d). Whereas control mice learned to perform the task progressively better over the course of 5–6 days, MdV-DTR mice fell off the rotating rod rapidly without significantly improved performance over 6 days (Fig. 4d, e). A similar lack of improved performance was also observed by applying a pharmacogenetic approach to transiently silence vGlut2^{ON} MdV neurons through expression of a PSEM[308] inducible silencer (PSAM$^{L141F\ Y115F}$–GlyR chimaeric ion channel; daily PSEM[308] administration 10 min before training)[37,38] (MdV-PSAM-GlyR mice; Fig. 4d, e). The same group of mice subsequently following a training regime without PSEM[308] administration picked up performance at a rate comparable to the control group (Fig. 4f), indicating that acute neuronal silencing is sufficient to explain poor task performance.

Second, we conducted a forelimb-specific skilled motor task (single-pellet reaching task), during which mice are placed in a box and learn to reach for single small food pellets placed outside a narrow slit and transfer them to their mouth[39,40] (Fig. 5a). We found that control mice learn to carry out this task with a success rate of ~45% (Fig. 5b), in agreement with previous findings[39]. In contrast, MdV-DTR mice were severely compromised in task performance, not improving over time and only completing successfully ~10% of all attempts even after 8 training days (Fig. 5b). We next analysed the different phases of this complex task (Supplementary Video 4) to determine whether MdV-DTR

©2014 Macmillan Publishers Limited. All rights reserved

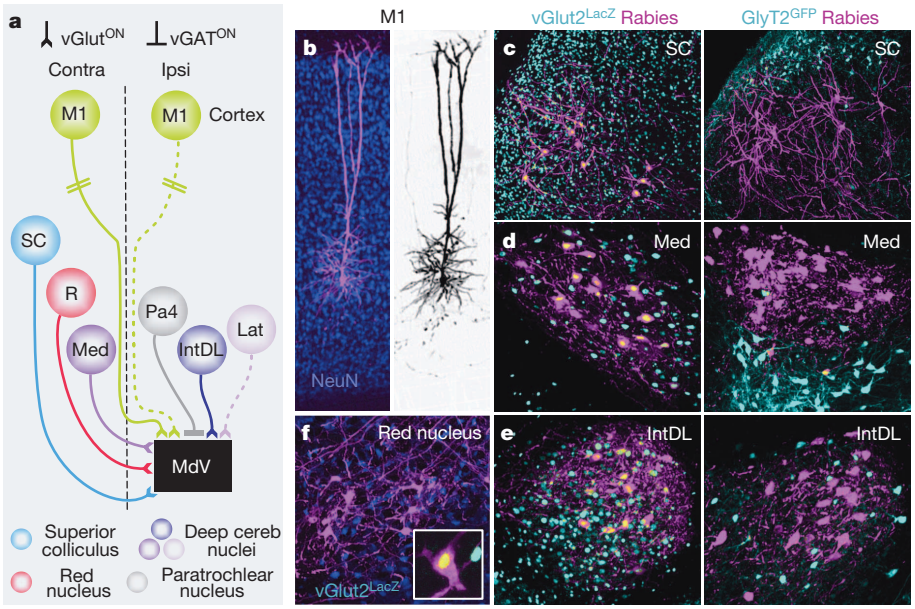


Figure 3 | MdV regulated by upstream motor centres. **a**, Diagram of brain areas with synaptic input to vGlut2^{ON} MdV neurons (black dashed line, midline; different arrows, excitatory and inhibitory connections; dashed lines, less prominent input to MdV than solid lines; $n = 6$ independent experiments). **b**–**f**, Neurons in primary motor cortex (M1, **b**), superior colliculus (SC, **c**), medial cerebellar nucleus (Med, **d**), interposed cerebellar nucleus, dorsolateral hump (IntDL, **e**) and red nucleus (**f**) connecting to vGlut2^{ON} MdV neurons. For **c**–**f** neurons are glutamatergic, marked by expression of vGlut2 (**c**–**f**) and absence of GlyT2 (**c**–**e**).

mice show specific defects. We found that both control and MdV-DTR mice learned to decrease the miss-rate during the initial reaching phase towards the pellet (Fig. 5c–e) and analysis of reaching trajectories, speed and variability revealed no difference between the two data sets (Fig. 5d). However, MdV-DTR mice had specific defects in fine paw placement on top of the pellet and in the closure of the fingers to correctly embrace the pellet for retrieval, resulting in increased error rates in the grasping category of the task compared to control mice (Fig. 5c, e). However, in cases where the pellet was grabbed correctly, MdV-DTR and control mice had similar failure rates in dropping the pellet during retrieval (Fig. 5e). Similar results were also obtained in MdV-PSAM-GlyR mice (Fig. 5f, g). To assess whether these deficiencies can be attributed to a general lack of descending drive to forelimb motor neurons, we performed force measurements of forelimb grip strength using previously described assays[41,42]. We found no reduction in forelimb grip strength in MdV-DTR mice compared to control mice or MdV-PSAM-GlyR mice before or after PSEM^{308} administration (Fig. 5e, g). Moreover, ablation of vGlut2^{ON} neurons in the neighbouring magnocellular reticular nucleus did not affect single-pellet reaching task performance or grip strength (Extended Data Fig. 10).

To determine whether MdV neurons are also needed for task execution after task learning, we ablated or acutely silenced vGlut2^{ON} MdV neurons only once mice achieved maximal success rates in the single-pellet reaching task (Fig. 6a, b). We found that as pre-learning inactivation experiments, post-learning MdV inactivation also significantly degraded motor performance specifically during the grasp phase (Fig. 6c, d). Together, these findings demonstrate that glutamatergic MdV neurons are required during specific movement phases to efficiently perform the single-pellet reaching task, both before and after learning the task.

Discussion

Descending pathways carry motor commands to the spinal cord, linking motor plan to task execution. Our study demonstrates that a larger variety of brainstem nuclei target forelimb than hindlimb motor neurons,

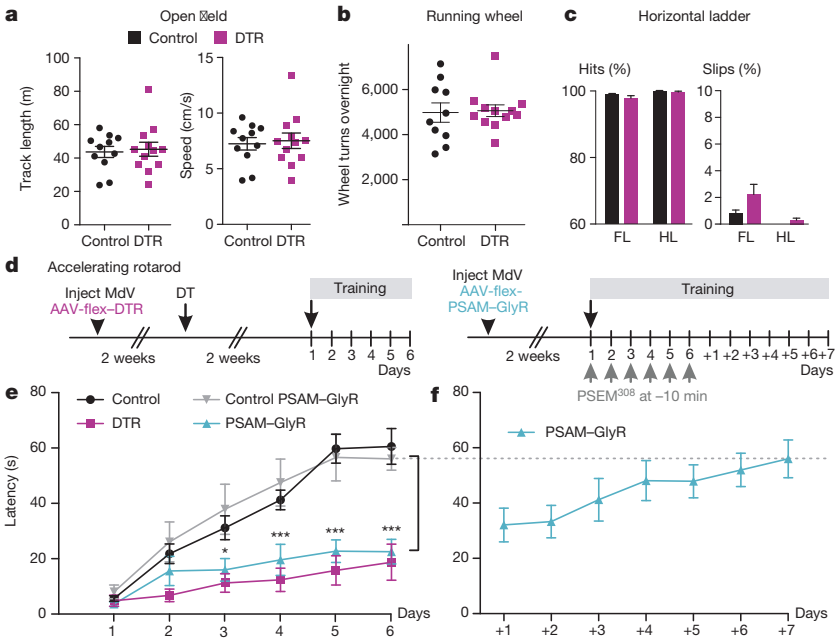


Figure 4 | Role of MdV neurons in motor behaviour. **a**–**c**, Performance analysis in basic motor tasks for control and MdV-DTR mice. Open field (**a**), running wheel (**b**, number of wheel turns measured over night) and horizontal ladder (**c**) are shown. In **a**, **b**, dots represent individual mice; for **c**, FL and HL analysed separately, $n = 10$ control and $n = 8$ MdV-DTR mice. **d**–**f**, Performance of control and MdV-DTR or MdV-PSAM-GlyR mice on the accelerating rotarod task (DT, diphtheria toxin). Time line for virus injections and task performance shown in **d** and latency quantification on rotarod on each day tested displayed in **e**, **f**. Number of mice analysed: control groups, $n = 14$; MdV-DTR, $n = 10$; PSAM-GlyR, $n = 12$ mice.

©2014 Macmillan Publishers Limited. All rights reserved

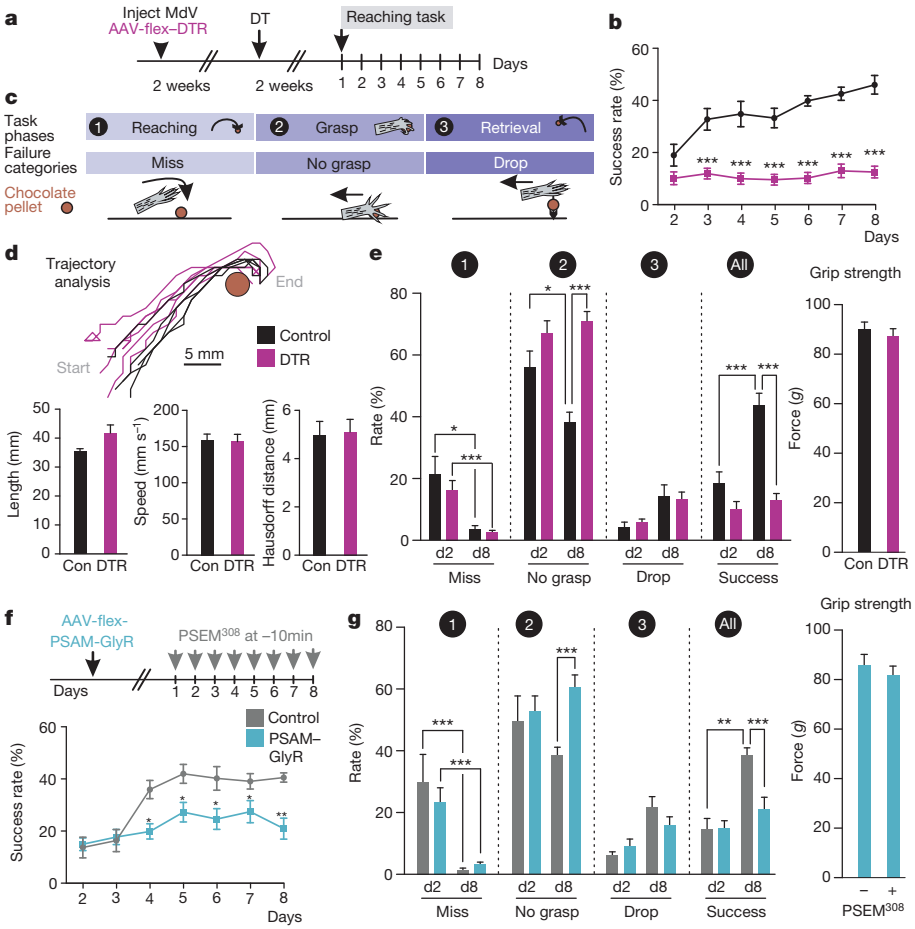


Figure 5 | MdV required for single-pellet reaching task. **a**, Time line for reaching task with ablation of glutamatergic MdV neurons. **b**, Success rate for reaching task of control ($n = 12$) and MdV-DTR ($n = 16$) mice from day 2 to 8. **c**, Representation of task phases (reaching, grasp, retrieval) and failure categories (miss, no grasp, drop). **d**, Trajectory analysis of successful reaching events in control and MdV-DTR mice. Average trajectory of ≥ 6 successful reaches from lifting paw off the ground to pellet contact (trajectory length, reaching speed and trajectory variability of averaged traces). **e**, Analysis of task phases for control and MdV-DTR mice on experimental day 2 and 8. Right plot, forelimb grip strength analysis for control and MdV-DTR mice ($n = 15$ mice for each). **f**, Time line and success rate for reaching task for MdV-PSAM-GlyR mice ($n = 9$) and control group ($n = 7$). **g**, Task phase analysis for control and MdV-PSAM-GlyR mice on experimental day 2 and 8. Right plot, forelimb grip strength analysis for MdV-PSAM-GlyR mice without or with PSEM^{308} administration ($n = 10$ mice).

correlating with the higher complexity and repertoire of forelimb motor tasks. We demonstrate the involvement of the forelimb-dominated brainstem nucleus MdV in expert performance of a skilled motor subroutine, supporting a model in which distinct brainstem subpopulations control aspects of motor behaviour through their specific targeted spinal subcircuits.

Forelimb-dominated brainstem nuclei show highly selective synaptic interaction matrices with spinal subcircuits, including specific motor neuron pools and spinal interneurons. Our findings agree with the concept that descending pathways to motor neurons operate through a dichotomous circuit connection strategy of both direct and indirect pathways. Direct pathways deliver commands to specific executing motor neuron pools, thus creating bypass circuits that may provide means to disregard the contribution of local spinal circuits often engaged in rhythmic activities[2,3,43,44]. In contrast, indirect motor command pathways contribute to local spinal computation by intersection with spinal microcircuits to enhance or suppress output channels to targeted motor neuron pools[14–17]. Dichotomy in descending circuit architecture and its possible correlation to sophistication of skilled movement has been most prominently discussed for motor cortical connections to spinal circuits[5–7]. Especially during challenging forelimb movements such as reaching or object manipulation, electromyographic recordings provide evidence for complex sequences of muscle contractions[45,46], and flexible divergent circuit architecture with access to specific motor neuron pools may allow achieving a higher degree of muscle synergies, known to be regulated by descending pathways[45,47]. Our study provides important

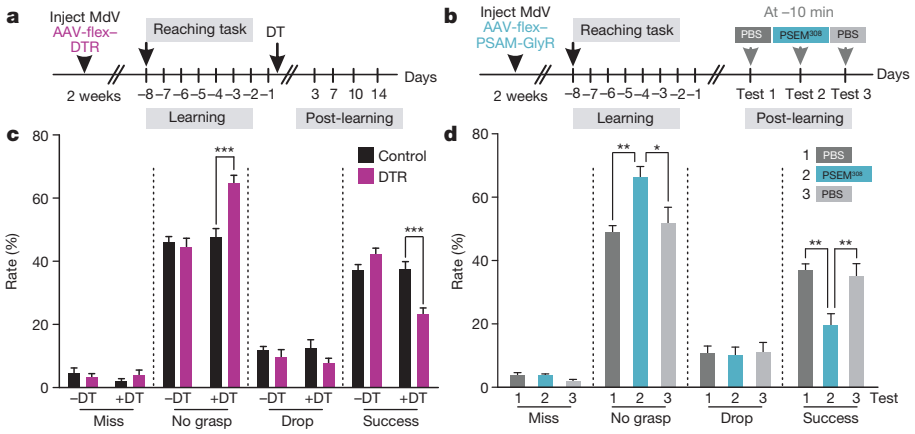


Figure 6 | MdV required for task execution. **a**, **b**, Time line for reaching task with ablation or silencing of glutamatergic MdV neurons after completion of learning. **c**, **d**, Task phase analysis for control ($n = 13$) and MdV-DTR ($n = 13$) mice with or without diphtheria toxin injection after completion of learning (**c**; 14 days post learning) and analogous analysis for MdV-PSAM-GlyR mice ($n = 8$) in three sequential post-learning tests (**d**; grey, PBS injection; cyan, PSEM^{308} injection).

©2014 Macmillan Publishers Limited. All rights reserved

insight into circuit-level mechanisms and anatomical substrates essential to implement task-specific motor subroutines during the execution of a complex motor program.

METHODS SUMMARY

AAV-G protein and trans-synaptic rabies viruses were produced and injected as described before[19,20,22,23]. For central nervous system-initiated trans-synaptic rabies experiments, local co-injection of AAVs conditionally expressing G protein and TVA was followed by injection of EnvA-coated rabies–mCherry[19,20,29]. Digital three-dimensional brainstem reconstructions were carried out by combinatorial use of Imaris (Bitplane) and Amira for neurons located caudal to the isthmus and rostral to decussation. The single-pellet reaching task was carried out following previously established procedures with slight modifications[39]. Other behavioural assays and experimental details are described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 July 2013; accepted 13 January 2014.
Published online 2 February 2014.

1. Shik, M. L. & Orlovsky, G. N. Neurophysiology of locomotor automatism. *Physiol. Rev.* **56,** 465–501 (1976).
2. Orlovsky, G. N., Deliagina, T. G. & Grillner, S. *Neuronal Control of Locomotion: From Mollusc to Man* (Oxford Univ. Press, 1999).
3. Grillner, S. Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron* **52,** 751–766 (2006).
4. Jordan, L. M., Liu, J., Hedlund, P. B., Akay, T. & Pearson, K. G. Descending command systems for the initiation of locomotion in mammals. *Brain Res. Rev.* **57,** 183–191 (2008).
5. Kuypers, H. G. J. M. in *Handbook of Physiology: The Nervous System. Anatomy of the Descending Pathways* 597–666 (American Physiological Society, 1981).
6. Lemon, R. N. Descending pathways in motor control. *Annu. Rev. Neurosci.* **31,** 195–218 (2008).
7. Rathelot, J. A. & Strick, P. L. Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *Proc. Natl Acad. Sci. USA* **106,** 918–923 (2009).
8. Alstermark, B. & Isa, T. Circuits for skilled reaching and grasping. *Annu. Rev. Neurosci.* **35,** 559–578 (2012).
9. Iwaniuk, A. N. & Whishaw, I. Q. On the origin of skilled forelimb movements. *Trends Neurosci.* **23,** 372–376 (2000).
10. Mori, S., Sakamoto, T., Ohta, Y., Takakusaki, K. & Matsuyama, K. Site-specific postural and locomotor changes evoked in awake, freely moving intact cats by stimulating the brainstem. *Brain Res.* **505,** 66–74 (1989).
11. Noga, B. R., Kriellaars, D. J., Brownstone, R. M. & Jordan, L. M. Mechanism for activation of locomotor centers in the spinal cord by stimulation of the mesencephalic locomotor region. *J. Neurophysiol.* **90,** 1464–1478 (2003).
12. Skinner, R. D. & Garcia-Rill, E. The mesencephalic locomotor region (MLR) in the rat. *Brain Res.* **323,** 385–389 (1984).
13. Garcia-Rill, E. & Skinner, R. D. The mesencephalic locomotor region. I. Activation of a medullary projection site. *Brain Res.* **411,** 1–12 (1987).
14. Drew, T., Prentice, S. & Schepens, B. Cortical and brainstem control of locomotion. *Prog. Brain Res.* **143,** 251–261 (2004).
15. Holmqvist, B. & Lundberg, A. Differential supraspinal control of synaptic actions evoked by volleys in the flexion reflex afferents in alpha motoneurones. *Acta Physiol. Scand. Suppl.* **186,** 1–15 (1961).
16. Grillner, S. & Hongo, T. Vestibulospinal effects on motoneurones and interneurones in the lumbosacral cord. *Prog. Brain Res.* **37,** 243–262 (1972).
17. Jankowska, E., Hammar, I., Slawinska, U., Maleszak, K. & Edgley, S. A. Neuronal basis of crossed actions from the reticular formation on feline hindlimb motoneurons. *J. Neurosci.* **23,** 1867–1878 (2003).
18. Jankowska, E. & Edgley, S. A. How can corticospinal tract neurons contribute to ipsilateral movements? A question with implications for recovery of motor functions. *Neuroscientist* **12,** 67–79 (2006).
19. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53,** 639–647 (2007).
20. Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nature Protocols* **5,** 595–606 (2010).
21. Ugolini, G. Advances in viral transneuronal tracing. *J. Neurosci. Methods* **194,** 2–20 (2010).
22. Stepien, A. E., Tripodi, M. & Arber, S. Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* **68,** 456–472 (2010).
23. Tripodi, M., Stepien, A. E. & Arber, S. Motor antagonism exposed by spatial segregation and timing of neurogenesis. *Nature* **479,** 61–66 (2011).
24. Paxinos, G. & Franklin, K. B. *The Mouse Brain in Stereotaxic Coordinates* 4th edn (Elsevier, 2012).
25. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71,** 142–154 (2011).
26. Pecho-Vrieseling, E., Sigrist, M., Yoshida, Y., Jessell, T. M. & Arber, S. Specificity of sensory-motor connections encoded by Sema3e–Plxnd1 recognition. *Nature* **459,** 842–846 (2009).
27. McKenna, J. E., Prusky, G. T. & Whishaw, I. Q. Cervical motoneuron topography reflects the proximodistal organization of muscles and movements of the rat forelimb: a retrograde carbocyanine dye analysis. *J. Comp. Neurol.* **419,** 286–296 (2000).
28. Greene, E. C. *Anatomy of the Rat* (Hafner, 1935).
29. Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M. & Callaway, E. M. Monosynaptic circuit tracing *in vivo* through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl Acad. Sci. USA* **107,** 21848–21853 (2010).
30. Gorski, J. A. *et al.* Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22,** 6309–6314 (2002).
31. Buitrago, M. M., Schulz, J. B., Dichgans, J. & Luft, A. R. Short and long-term motor skill learning in an accelerated rotarod training paradigm. *Neurobiol. Learn. Mem.* **81,** 211–216 (2004).
32. Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462,** 920–924 (2009).
33. Costa, R. M., Cohen, D. & Nicolelis, M. A. Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr. Biol.* **14,** 1124–1134 (2004).
34. Bretzner, F. & Brownstone, R. M. Lhx3-Chx10 reticulospinal neurons in locomotor circuits. *J. Neurosci.* **33,** 14681–14692 (2013).
35. Buch, T. *et al.* A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nature Methods* **2,** 419–426 (2005).
36. Ruediger, S. *et al.* Learning-related feedforward inhibitory connectivity growth required for memory precision. *Nature* **473,** 514–518 (2011).
37. Atasoy, D., Betley, J. N., Su, H. H. & Sternson, S. M. Deconstruction of a neural circuit for hunger. *Nature* **488,** 172–177 (2012).
38. Magnus, C. J. *et al.* Chemical and genetic engineering of selective ion channel-ligand interactions. *Science* **333,** 1292–1296 (2011).
39. Xu, T. *et al.* Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462,** 915–919 (2009).
40. Whishaw, I. Q. & Pellis, S. M. The structure of skilled forelimb reaching in the rat: a proximally driven movement with a single distal rotatory component. *Behav. Brain Res.* **41,** 49–59 (1990).
41. Blanco, J. E., Anderson, K. D. & Steward, O. Recovery of forepaw gripping ability and reorganization of cortical motor control following cervical spinal cord injuries in mice. *Exp. Neurol.* **203,** 333–348 (2007).
42. Bui, T. V. *et al.* Circuits for grasping: spinal dI3 interneurons mediate cutaneous control of motor behavior. *Neuron* **78,** 191–204 (2013).
43. Kiehn, O. Development and functional organization of spinal locomotor circuits. *Curr. Opin. Neurobiol.* **21,** 100–109 (2011).
44. Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Rev. Neurosci.* **10,** 507–518 (2009).
45. Yakovenko, S., Krouchev, N. & Drew, T. Sequential activation of motor cortical neurons contributes to intralimb coordination during reaching in the cat by modulating muscle synergies. *J. Neurophysiol.* **105,** 388–409 (2011).
46. Hyland, B. I. & Jordan, V. M. Muscle activity during forelimb reaching movements in rats. *Behav. Brain Res.* **85,** 175–186 (1997).
47. Roh, J., Cheung, V. C. & Bizzi, E. Modules in the brain stem and spinal cord underlying motor behaviors. *J. Neurophysiol.* **106,** 1363–1378 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to M. Mielich, B. Rapp, D. Salvador and M. Sigrist for expert technical help, L. Gelman, A. Ponti, N. Ehrenfeuchter, M. Kirschmann and R. Thierry for help and advice with image acquisition and analysis, and to P. Caroni for discussions and comments on the manuscript. We thank F. Donato, Y. Zuo, A. Takeoka, P. Tovote and P. Botta for advice on design of behavioral experiments, and S. Sternson for providing PSEM[308] and advice for neuronal silencing experiments. M.S.E. was supported by a long-term fellowship of the Human Frontier Science Program, and all authors by a European Research Council Advanced Grant, the Swiss National Science Foundation, the Kanton Basel-Stadt and the Novartis Research Foundation.

Author Contributions M.S.E. was involved in design of experiments, carried out experiments, acquired and analysed data. P.C. carried out experiments, acquired and analysed data. S.A. initiated the project, designed experiments, analysed data and wrote the manuscript. All authors discussed the experiments and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A. (silvia.arber@unibas.ch).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13021

Skilled reaching relies on a V2a propriospinal internal copy circuit

Eiman Azim[1], Juan Jiang[2], Bror Alstermark[2] & Thomas M. Jessell[1]

The precision of skilled forelimb movement has long been presumed to rely on rapid feedback corrections triggered by internally directed copies of outgoing motor commands, but the functional relevance of inferred internal copy circuits has remained unclear. One class of spinal interneurons implicated in the control of mammalian forelimb movement, cervical propriospinal neurons (PNs), has the potential to convey an internal copy of premotor signals through dual innervation of forelimb–innervating motor neurons and precerebellar neurons of the lateral reticular nucleus. Here we examine whether the PN internal copy pathway functions in the control of goal–directed reaching. In mice, PNs include a genetically accessible subpopulation of cervical V2a interneurons, and their targeted ablation perturbs reaching while leaving intact other elements of forelimb movement. Moreover, optogenetic activation of the PN internal copy branch recruits a rapid cerebellar feedback loop that modulates forelimb motor neuron activity and severely disrupts reaching kinematics. Our findings implicate V2a PNs as the focus of an internal copy pathway assigned to the rapid updating of motor output during reaching behaviour.

Skilled forelimb movements constitute some of the more impressive accomplishments of the mammalian motor system[1–3]. Goal-directed reaching involves the activation of descending pathways that provide commands for task-appropriate motor programs[4–6]. Less clear is the issue of how such descending commands engage spinal circuits to achieve the modularity and precision evident in reach, grasp and object manipulation. One view holds that skilled motor performance requires on-line refinement throughout the course of movement[7–9], through internally directed copies of motor commands that engage cerebellar circuits and permit rapid updating of motor output[9–14]. But putative internal copy pathways, by their nature, are closely interwoven with motor output circuits, a feature that has made it hard to isolate the neural substrate of such internal copies or to assess whether they do, in fact, influence motor performance.

One interneuron class, cervical PNs, has long been implicated in the control of forelimb behaviour[15,16]. In cat and primate, PNs comprise excitatory and inhibitory neuronal subtypes that serve as intermediary relays for descending motor commands[16,17]. PNs are characterized by an ipsilateral bifurcated output: one axonal branch projects caudally to the cervical motor neurons that control forelimb muscles[18,19], and the other projects rostrally to the lateral reticular nucleus (LRN)[20], a precerebellar relay[21–24] (Fig. 1a). In principle, the intriguing duality of PN axonal projections offers a simple anatomical substrate for the internal copying of premotor signals. In cat, severing the premotor axonal branch of PNs by lesioning the ventrolateral funiculus perturbs reaching but not grasping[25], whereas silencing PN output in monkey perturbs both reaching and grasping[26]. Neither of these manipulations, however, has addressed the relevance of an internal copy branch for on-line refinement of motor output.

In this study, we sought to define the contribution of excitatory PNs to skilled reach behaviour in mice. We reasoned that their molecular delineation could provide a genetic means of eliminating PNs as well as manipulating their internal copy projections. We show that one major population of excitatory PNs belongs to the Chx10 (also known as Vsx2)-expressing V2a interneuron class[27–29]—one of the cardinal subtypes of ventral interneurons implicated in motor control[30]. Genetic elimination of cervical V2a interneurons elicits a reach-specific defect in forelimb movement, revealed by quantitative kinematics. Selective activation of the PN internal copy branch triggers a rapid cerebellar feedback loop that excites motor neurons and degrades forelimb movement. Our findings show that excitatory PNs establish an internal feedback circuit assigned to the control of mammalian skilled reaching.

Organization of PNs in mouse

To explore the existence and organization of excitatory PNs in mice we focused on two defining features: a bifurcating output to the LRN and forelimb-innervating motor neurons, and receipt of a direct input from reticulospinal neurons (Fig. 1a and Supplementary Note 1)[16,17,19,20].

Extracellular recording from cervical spinal segments (C3, C4, C6) in adult anaesthetized mice identified neurons that could be activated antidromically by microstimulation of the ipsilateral LRN as well as C7 segment ventral horn (Fig. 1b). Spike collision tests confirmed that individual neurons send axons to both targets (Fig. 1b; red arrowheads). To investigate the premotor output of PNs we recorded intracellularly from forelimb-innervating motor neurons. Stimulation of the LRN with the intent of antidromically activating PNs produced short-latency monosynaptic excitatory postsynaptic potentials (EPSPs) in motor neurons (Fig. 1c; mean latency from stimulation 1.24 (from volley 0.54) $\pm$ 0.13 ms s.d.; mean within-cell variance of EPSP onset 0.008 $\pm$ 0.006 ms^2 s.d.). LRN simulation elicited pure EPSPs in 76%, pure inhibitory postsynaptic potentials (IPSPs) in 16% and mixed responses in 8% of forelimb-innervating motor neurons ($n = 38$ neurons), indicating the predominance of excitatory PNs. Stimulation of the reticulospinal tract elicited short-latency monosynaptic motor neuron EPSPs[31] that summed with EPSPs induced by LRN stimulation (Fig. 1c; mean latency from stimulation 1.18 (from volley 0.68) $\pm$ 0.16 ms s.d.), indicating independent recruitment of PN and reticulospinal inputs (Supplementary Note 2).

We examined whether reticulospinal neurons provide direct input to PNs. As above, extracellular recording was used to identify PNs

[1]Howard Hughes Medical Institute, Kavli Institute for Brain Science, Mortimer B. Zuckerman Mind Brain Behavior Institute, Departments of Neuroscience and Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. [2]Department of Integrative Medical Biology, Section of Physiology, Umeå University, Umeå, Sweden.

©2014 Macmillan Publishers Limited. All rights reserved

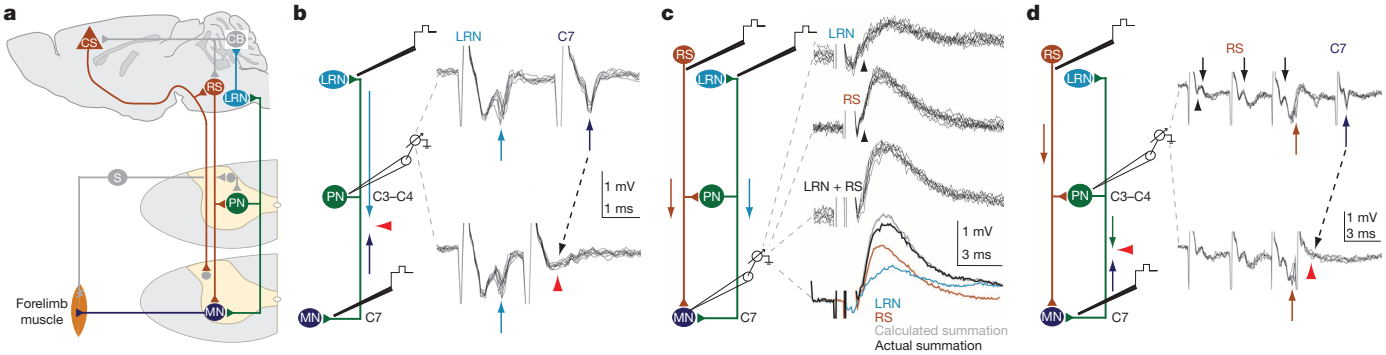


Figure 1 | Identification of mouse PNs. **a**, PNs receive direct and indirect supraspinal (corticospinal (CS), reticulospinal (RS)) and sensory (S) input. PNs innervate cervical motor neurons (MN) and LRN neurons that project to cerebellum (CB). **b**, *In vivo* extracellular recordings (C3–C4) while stimulating LRN (20–100 μA) and C7 ventral horn (40 μA) revealed antidromic spikes (arrows) and collision (red arrowheads; $n = 12$). **c**, Intracellular motor neuron recordings during LRN stimulation revealed monosynaptic EPSPs (100 μA; $n = 29$). LRN-induced EPSPs summate with monosynaptic EPSPs elicited by reticulospinal stimulation (100 μA; $n = 34$; arrowheads, EPSP onset). **d**, Extracellular PN recordings, identified via C7 (17 μA; blue arrows) and LRN stimulation (not shown), revealed monosynaptic spikes following reticulospinal stimulation (3 × 50 μA; red arrows; $n = 14$) and collision (red arrowheads; $n = 17$). See Supplementary Note 1.

through antidromic stimulation from both LRN and C7. Stimulation of the reticulospinal tract elicited short-latency monosynaptic responses (Fig. 1d; mean latency from third volley 0.43 ± 0.17 ms s.d.; collision, red arrowheads). Taken together, these findings indicate that key features of PN circuitry are conserved in mice (Supplementary Note 1)[16].

Excitatory PNs are a V2a subpopulation

V2a interneurons represent a major class of ipsilateral, excitatory premotor interneurons in the ventral spinal cord, and express the homeodomain transcription factor Chx10 (Fig. 2a)[27,28]. To assess whether excitatory PNs are included within the V2a cohort we crossed *Chx10-cre* mice with a *Rosa-lsl-tdTomato* reporter line, generating *Chx10::tdT* mice. V2a interneurons were labelled by unilateral injection of a Cre-dependent (flip-excision, FLEX) adeno-associated viral vector expressing a channelrhodopsin 2–yellow fluorescent protein fusion (*AAV-FLEX-hChR2-YFP*) into rostral cervical segments of adult *Chx10::tdT* mice (Fig. 2a).

With this labelling strategy >80% of all rostral cervical tdT$^+$ V2a interneurons selectively expressed YFP (Fig. 2b). The LRN contained a dense network of YFP$^+$ axons, and individual LRN neurons were studded with vesicular glutamate transporter 2 (vGluT2, also known as Slc17a6)-expressing, YFP$^+$ V2a-derived boutons (Fig. 2c). Similarly, C7–C8-level choline acetyltransferase (ChAT)$^+$ motor neurons were contacted by vGluT2$^+$, YFP$^+$ V2a boutons (Fig. 2d). We detected a similar pattern of bouton labelling on LRN and motor neurons after viral injection at C6–thoracic (T)1 levels, but little YFP$^+$ labelling in the LRN after viral injection at more caudal levels (Extended Data Fig. 1a and Supplementary Note 3)[20]. Thus V2a neurons within C3 to T1 segments provide glutamatergic input to LRN and motor neurons.

To assess whether individual cervical V2a interneuron axons bifurcate to innervate LRN and forelimb-innervating motor neuron targets we injected a retrograde tracer, cholera toxin subunit B (CTB), into the LRN and a trans-synaptic tracer, wheat germ agglutinin (WGA)[32], into forelimb muscles of *Chx10-cre;Rosa-lsl-YFP* mice. More than 30% of YFP$^+$ cervical V2a interneurons were CTB-labelled and many YFP$^+$,

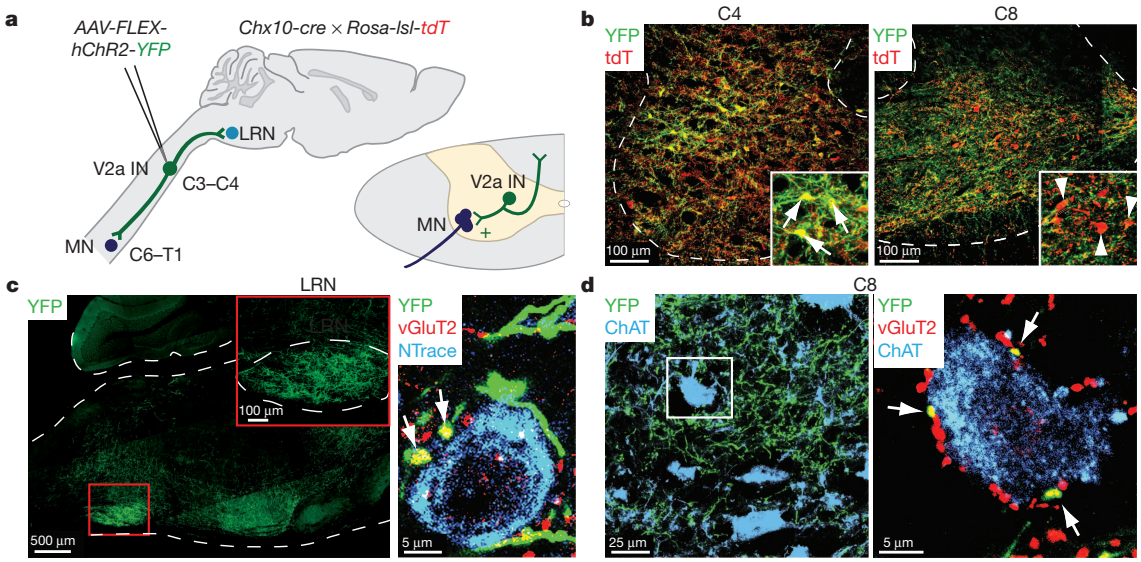


Figure 2 | Excitatory PNs are a V2a interneuron subpopulation. **a**, Unilateral C3–C4 injection of *AAV-FLEX-hChR2-YFP* into *Chx10::tdT* mice. IN, interneuron. **b**, 82% (±7% s.e.m.; $n = 2$) of C3–C4 tdT$^+$ V2a interneurons were transduced (C4, yellow neurons, arrows; C8, red neurons, arrowheads). **c**, YFP$^+$ V2a interneurons project to LRN. Neurons, labelled with NeuroTrace (NTrace, blue), are studded with vGluT2$^+$ (red), YFP$^+$ boutons (arrows, yellow boutons). Sparse YFP$^+$ axonal labelling is also found in facial nucleus. **d**, C7–C8 ChAT$^+$ motor neurons (blue) are contacted by vGluT2$^+$ (red), YFP$^+$ boutons (arrows, yellow boutons).

©2014 Macmillan Publishers Limited. All rights reserved

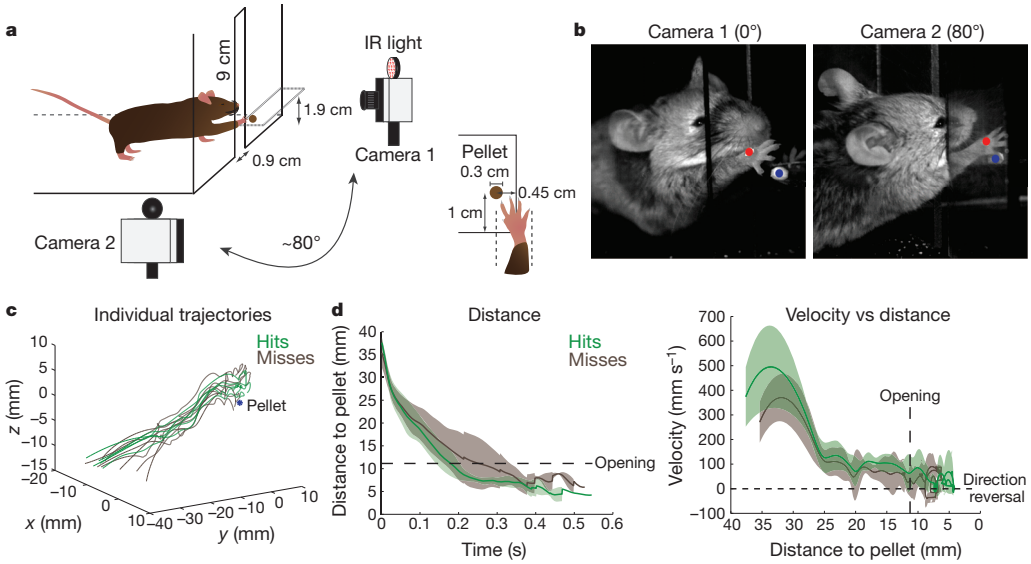


Figure 3 | Reaching kinematics. **a**, Mice were trained to reach for a food pellet displaced to the left. IR, infrared. **b**, High-speed capture and tracking of an infrared-reflective marker attached to the right paw. **c**, Trajectories of successful (hits, green traces) and unsuccessful (misses, brown traces) trials from a representative mouse. **d**, Mean kinematics from a representative mouse. Transition from reach to grab phase delineated by box opening (large dashes). Velocity crossings of zero (small dashes) indicate direction reversals. Shaded regions indicate s.d. See Extended Data Fig. 2b.

CTB^+, WGA^+ neurons were detected at intermediate levels of cervical spinal cord (Extended Data Fig. 1b). Thus a substantial fraction of cervical V2a interneurons are excitatory PNs.

Kinematics of mouse reaching

To evaluate the behavioural role of PNs we developed a three-dimensional kinematic assay of paw position during a goal-directed reaching task (Fig. 3a, b, Supplementary Videos 1, 2 and Methods)[33,34]. We delineated three modular aspects of movement: an early proximal 'reach' phase, from forelimb lift to the point that the paw passes through the access window; a late distal 'grab' phase, in which the arm is fully extended and the paw pronates in anticipation of pellet retrieval; and a digit abduction phase, during grasp. Task success, defined by pellet retrieval, was assessed in a multi-reach assay that tolerates numerous reach attempts[34], and a more stringent single-trial kinematic reach assay (Supplementary Note 4). In both assays the incidence of trial success varied considerably between animals (Extended Data Fig. 2a)[33], leading us to compare behaviour within individual mice, before and after experimental perturbation.

Analysis of paw movement in wild-type mice revealed stereotypic reach kinematics: a comparison of 14 reach parameters failed to detect systematic differences between successful and unsuccessful trials (Fig. 3c, d, Extended Data Fig. 2b, Supplementary Videos 1, 2 and Methods; $n = 7$; two-tailed paired t-test), implying that wild-type trial outcome reflects the accuracy of digit placement and/or object manipulation rather than an overt deviation in paw kinematics. We used this behavioural assay to explore two genetic means of perturbing V2a PN function: acute ablation of cervical V2a neurons, eliminating both PN axonal outputs; and optogenetic stimulation of PN terminals in the LRN to perturb selectively the internal copy branch.

V2a interneuron ablation perturbs reaching

We targeted cervical V2a interneurons for elimination by unilateral injection of a viral vector that directed conditional expression of a diphtheria toxin receptor (DTR)–GFP fusion (*AAV-FLEX-DTR-GFP*) into C3–T1 levels of adult *Chx10::tdT* mice (Fig. 4a and Extended Data Fig. 3a)[35]. Seven days after viral injection >80% of tdT^+ V2a interneurons selectively expressed DTR–GFP, and the LRN contained a dense network of GFP^+ axons (Fig. 4b and Extended Data Fig. 3b). Diphtheria toxin (DT; 400 ng) administration 14–21 days after DTR transduction resulted, 7 days later, in a >80% elimination of C3–T1 tdT^+ V2a interneurons and a virtually complete loss of GFP^+ axons within the LRN (Fig. 4c and Extended Data Fig. 3b, c).

After cervical V2a interneuron ablation, reach success was reduced in the multi-reach task and eliminated in the kinematic assay, with no impact in DT-treated control mice spared DTR transduction (Fig. 4d and Extended Data Fig. 3d). During the reach phase in V2a-interneuron-depleted, but not in control, mice we observed an ~sixfold increase in the incidence of paw direction reversals, an ~twofold decrease in mean velocity and an ~twofold increase in the total duration of movement (Fig. 4e, f, Extended Data Fig. 3e–g, Extended Data Table 1 and Supplementary Video 3). Motor impairment was confined to the reach phase, as the grab phase did not exhibit perturbed kinematics (Fig. 4e, Extended Data Fig. 3e and Extended Data Table 1). Moreover, the extent of digit abduction during grasp attempts was unaffected (Extended Data Fig. 3h and Supplementary Video 3), implying a sparing of digit movements. Thus the reach phase is perturbed selectively by ablation of cervical V2a interneurons.

To address the specificity of motor impairment mice were tested for accuracy of forepaw placement using a horizontal ladder task[36] and were found to display an equivalently low incidence of rung-placement errors before and after DT-mediated ablation (Extended Data Fig. 3i and Supplementary Note 5). Thus elimination of cervical V2a neurons exerts a preferential impact on forelimb reach behaviour.

PN internal copy branch activation

The reaching impairments observed after cervical V2a neuron elimination left unresolved the relevance of the PN internal copy branch. To address this issue ChR2 was expressed in cervical V2a interneurons by unilateral injection of *AAV-FLEX-hChR2-YFP* into C3–T1 segmental levels in adult *Chx10-cre* mice. We reasoned that focal laser photostimulation of $ChR2^+$ PN terminals within the LRN could provide specific access to the PN subset of V2a interneurons and permit selective manipulation of the PN internal copy pathway.

After viral injection we detected a dense network of ChR2–YFP^+ fibres within the LRN (Fig. 2 and Extended Data Fig. 1a). Extracellular recording from LRN neurons in anaesthetized mice revealed that focal photostimulation of PN terminals excited ~40% of neurons with cerebellar projections ($n = 8$ out of 21), assessed by spike collision after cerebellar stimulation, with no effect in virus-spared control mice (Fig. 5a and Extended Data Fig. 4a–c).

©2014 Macmillan Publishers Limited. All rights reserved

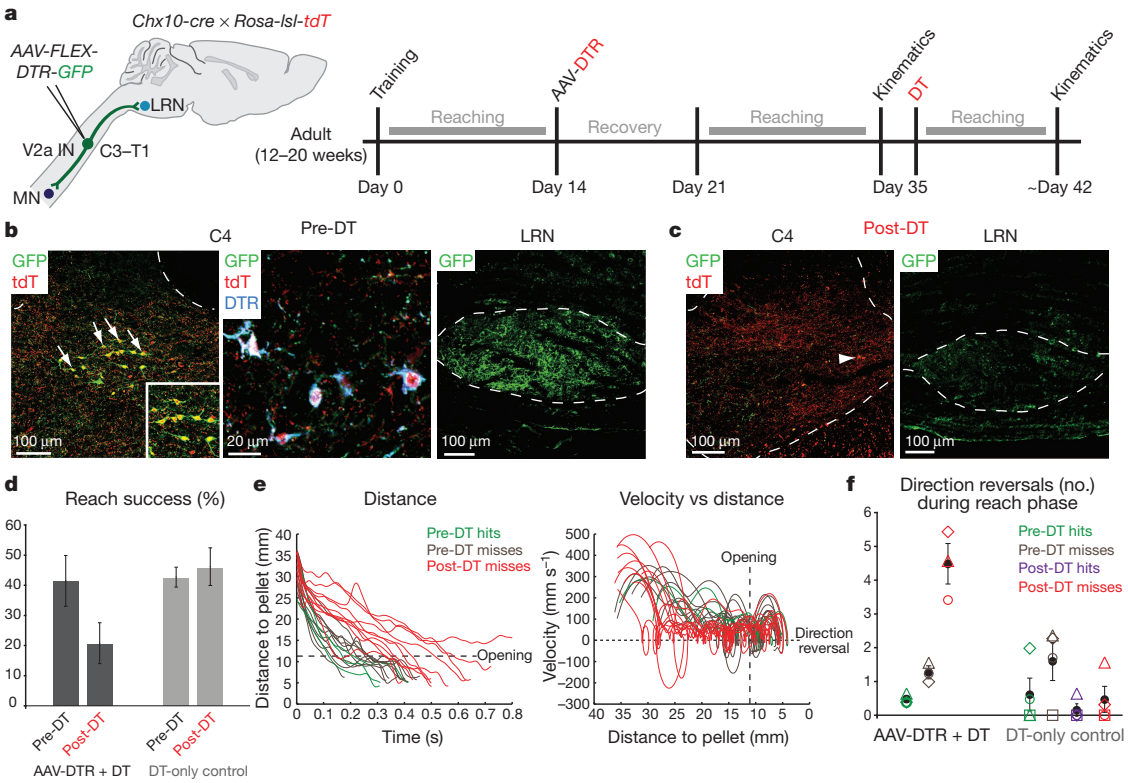


Figure 4 | Cervical V2a interneuron ablation perturbs reaching. **a**, Experimental design. **b**, After C3–T1 *AAV-FLEX-DTR-GFP* injection 83% (±0.3% s.e.m.; $n = 2$) of tdT$^+$ V2a interneurons co-expressed GFP and DTR (arrows), and GFP$^+$ projections were detected in LRN. **c**, Post-DT, 84% (±9% s.e.m.; $n = 2$) of C3–T1 tdT$^+$ V2a interneurons and GFP$^+$ LRN projections were eliminated (arrowhead, spared V2a interneuron). **d**, Ablation reduced success in the multi-reach task ($n = 3$ DTR, $n = 4$ control). **e**, Kinematics from a representative V2a interneuron-ablated mouse. See Extended Data Fig. 3e. **f**, Mean number of direction reversals increased during the reach, but not grab, phase in V2a interneuron-ablated mice. Shapes represent individual mice. Error bars indicate mean ± s.e.m. See Extended Data Table 1 and Extended Data Fig. 3 for statistical analysis.

We examined whether photostimulation of PN terminals in the LRN triggers antidromic action potentials. As above, PNs were identified by spike collision after electrical stimulation of LRN and C7 (Fig. 5b and Extended Data Fig. 5a). Photostimulation of identified PN cell bodies at cervical levels revealed that ~70% expressed ChR2, as determined by spike collision after LRN or C7 electrical stimulation (Fig. 5c, Extended Data Fig. 5b and Supplementary Note 6). For these PNs we compared the impact of local LRN electrical stimulation and focal LRN photostimulation, monitoring the incidence of antidromic spikes. Electrical stimulation of the LRN invariably elicited antidromic spikes in PNs, whereas LRN photostimulation never elicited spiking (0 out of 31 PNs; Fig. 5d and Extended Data Fig. 5c). Thus photostimulation of ChR2$^+$ PN terminals within the LRN activates the internally directed pathway without eliciting antidromic action potentials (Supplementary Discussion).

To examine the behavioural impact of activating the PN internal branch we analysed the influence of LRN photostimulation on reach kinematics (Fig. 6a). Application of light pulses severely degraded reach success in the multi-reach task and eliminated success in the kinematic

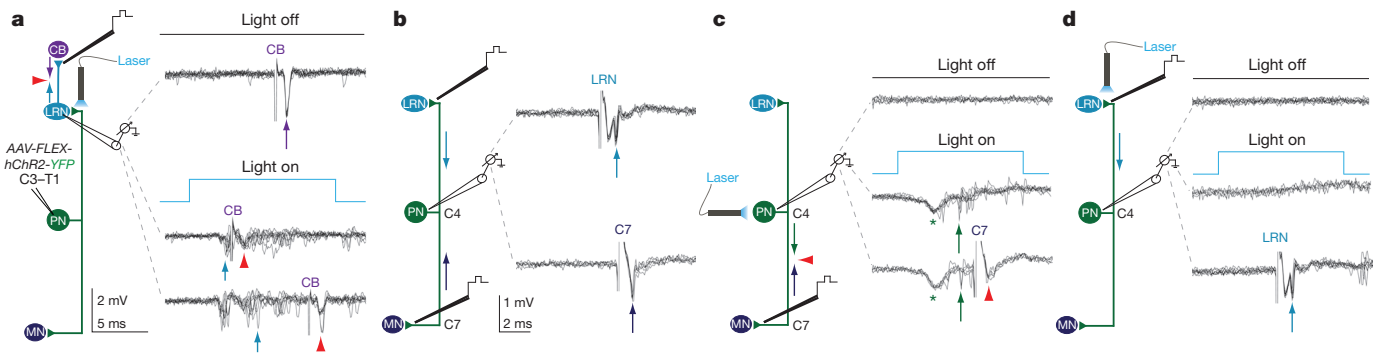


Figure 5 | Photoactivation of PN terminals in the LRN. **a**, Extracellular recording of LRN neurons antidromically activated from cerebellum (20 µA; purple arrows) during PN terminal photostimulation (473 nm, 20 Hz, ~12 mW) revealed repeated spiking ($n = 8$ out of 21; blue arrows) and collision (red arrowheads, two collision failures in bottom traces). **b**, PNs in C3–C4 were identified by electrical stimulation from LRN (80 µA) and C7 (40 µA) and by collision (not shown). **c**, Photostimulation of the same PN cell bodies activated 69% of PNs ($n = 9$ out of 13; green arrows), verified by collision of C7 spike (bottom traces, red arrowheads; compare with antidromic spike in **b**). See Supplementary Note 6. **d**, In the same PNs, photostimulation of terminals in LRN did not trigger antidromic spikes (0 out of 31 PNs; 0 out of 3 control), whereas electrical stimulation in LRN always produced antidromic spikes (bottom traces, arrow; compare with antidromic spike in **b**). See Extended Data Fig. 5 and Supplementary Discussion.

©2014 Macmillan Publishers Limited. All rights reserved

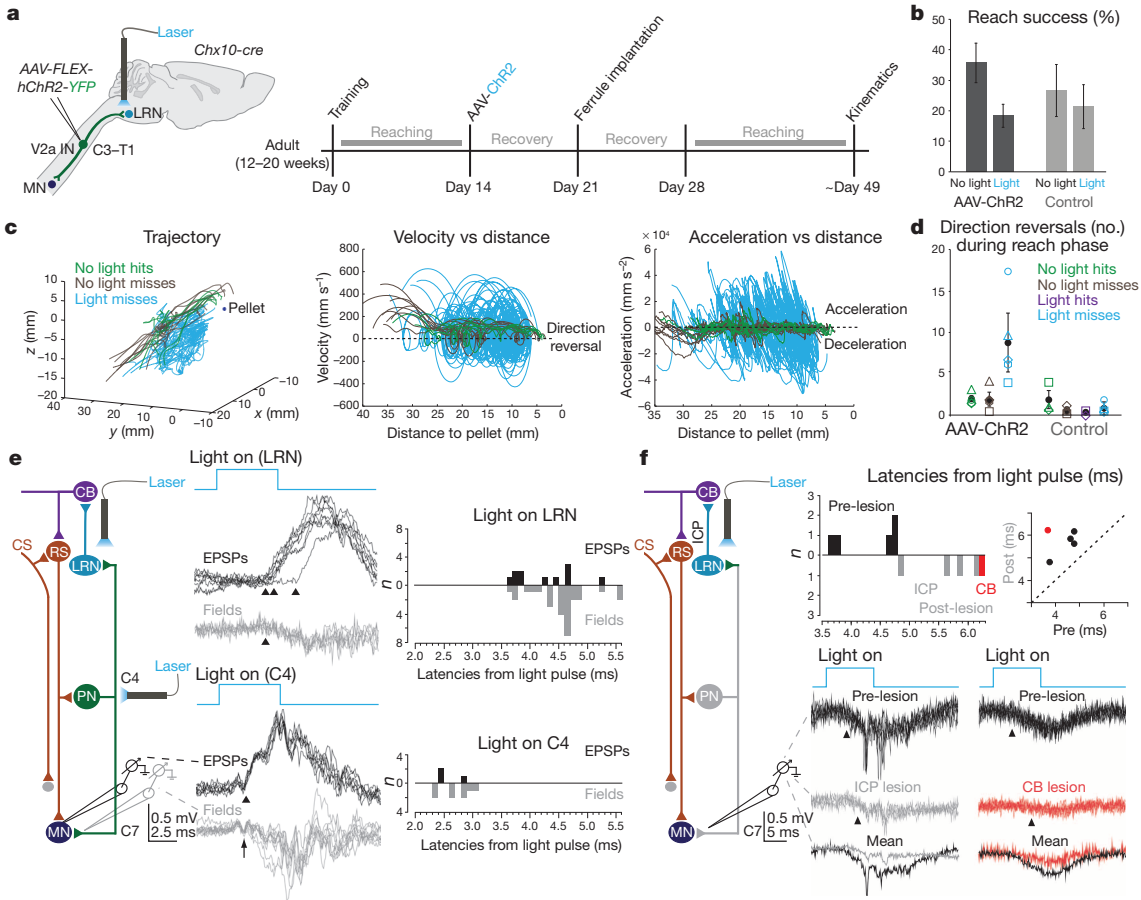


Figure 6 | PN terminal photostimulation perturbs forelimb movement via a cerebellar–motor loop. **a**, Experimental design. **b**, Photostimulation reduced success in the multi-reach task ($n = 5$ ChR2, $n = 4$ control). **c**, Kinematics from a representative AAV-ChR2 mouse. See Extended Data Fig. 6a. **d**, Mean number of reach phase direction reversals increased during photostimulation. See Extended Data Table 2 and Extended Data Fig. 6 for statistical analysis. **e**, Intracellular recording from forelimb motor neurons during PN terminal photostimulation revealed EPSPs with varying onset (top black traces, bars; 0.8 ms between first two arrowheads, 2.5 ms between first and third arrowheads; $n = 11$). Cervical photostimulation of PN cell bodies produced shorter latency fixed-onset EPSPs (bottom black traces, bars; onset (arrowhead) 0.6 ms from volley in field (arrow); $n = 3$). Field potentials recorded in C6–C7 ventral horn (grey traces, grey bars) had similar onset and duration as motor neuron EPSPs ($n = 27$, LRN-light; $n = 8$, C4-light). **f**, Bilateral lesion of inferior cerebellar peduncles (ICP; grey traces) or cerebellar extirpation (CB; red traces) reduced field potential size (mean reduction in area; 56% ± 8.8% s.d.; $P = 0.001$, two-tailed paired t-test; $n = 5$). Shortest latency fields (~3.6–4.7 ms; arrowheads, onset) were eliminated after lesions (histogram, scatter plot). Error bars indicate mean ± s.e.m.

assay, with no impact on control, virus-spared mice (Fig. 6b). Indeed, photostimulation in virally transduced, but not in control, mice severely perturbed motor performance (Fig. 6c, d, Extended Data Fig. 6a–c, Extended Data Table 2 and Supplementary Video 4, no light; Supplementary Video 5, light). We detected an ~4.5-fold increase in the incidence of paw direction reversals during the reach phase (Fig. 6d and Extended Data Table 2). Certain other kinematic parameters, notably variations in velocity, were perturbed in both reach and grab phases (Fig. 6c, Extended Data Fig. 6a, c and Extended Data Table 2)—nevertheless digit abduction was not affected (Extended Data Fig. 6d and Supplementary Video 5), revealing an element of selectivity. These findings indicate that imposed activation of the internally directed PN branch severely disrupts forelimb reaching movements (Supplementary Discussion).

A PN internal copy cerebellar–motor loop

A core feature of internal copy models is a rapid modulation of motor output[10–14]. We therefore explored how motor neurons are influenced by photostimulation of the internally directed PN branch. Intracellular recording from forelimb-innervating motor neurons in anaesthetized *Chx10-cre* mice injected at C3–T1 with *AAV-FLEX-hChR2-YFP* revealed polysynaptic EPSPs after LRN photostimulation (Fig. 6e, top black traces; mean onset latency 4.18 ± 0.75 ms s.d.; mean variance 0.28 ± 0.35 ms^2 s.d.). By contrast, recruitment of the premotor branch by photostimulation of PN cell bodies revealed monosynaptic motor neuron EPSPs (Fig. 6e, bottom black traces; mean onset latency 2.60 ± 0.20 ms s.d.; mean variance 0.014 ± 0.009 ms^2 s.d.). Recording from LRN neurons during PN terminal photostimulation revealed short-latency, monosynaptic activation (mean onset latency 0.87 ± 0.10 ms s.d. from local depolarization; mean variance 0.009 ± 0.006 ms^2 s.d.), evidence that the variability in motor neuron EPSP onset after PN terminal photostimulation arises downstream of the PN–LRN synapse (Extended Data Fig. 7a). These findings reinforce the view that photostimulation of PN terminals activates motor neurons through a polysynaptic pathway and not by antidromic recruitment of the premotor branch (Supplementary Discussion and Extended Data Fig. 7b).

To examine whether the PN–LRN internal copy pathway engages cerebellar circuits we severed LRN output to cerebellar cortex and deep cerebellar nuclei through bilateral transection of the inferior cerebellar peduncles, or by cerebellar extirpation (Extended Data Fig. 8a). These lesions did not impair synaptic activation of LRN neurons induced by PN terminal photostimulation (Extended Data Fig. 8b). Motor neuron responses were assessed by recording field potentials in C6–C7 ventral horn, revealing an onset latency and duration similar to that of motor neuron EPSPs (Fig. 6e, grey traces; Extended Data Fig. 8c and Supplementary Note 7). After cerebellar lesions, field potentials elicited by PN terminal photostimulation were reduced by up to ~65% (Fig. 6f), suggesting

©2014 Macmillan Publishers Limited. All rights reserved

that both cerebellar and non-cerebellar output pathways[37–39] mediate the PN–LRN control of motor output. Nevertheless, field potential onsets after lesions occurred with an additional delay of ~1–3 ms when compared with non-lesioned mice (Fig. 6f). Thus the cerebellar-directed PN internal copy pathway appears to be involved in rapid feedback modulation of motor output.

Discussion

The refinement of goal-directed reaching movements is thought to rely on the generation of internally directed copies of motor commands[9–16]. Here we provide evidence that V2a PNs lie at the core of an internal copy feedback circuit crucial for skilled reaching. Ablation of cervical V2a interneurons elicits selective reaching defects, photostimulation of PN terminals within the LRN erodes the fidelity of forelimb movement, and activation of the PN internal copy branch recruits a fast cerebellar–motor feedback loop.

Our functional analysis has provided insights into two general issues in limb motor control: the degree to which distinct microcircuits control elemental aspects of movement, and the evolutionary conservation of circuits and strategies for skilled reaching. Targeted elimination of PNs, albeit through ablation of the entire set of cervical V2a interneurons, disrupts forelimb reaching in a selective manner. Thus PNs in particular, and cervical V2a interneurons as a whole, appear to have little impact on aspects of mouse forelimb movement that engage distal musculature during grasping. These findings add to an emerging view that limb motor modularity has its basis in the recruitment of distinct spinal interneuron subtypes and circuits[1,16,25,40]. Genetic inactivation of dI3 excitatory premotor interneurons preferentially impairs forepaw grasp behaviours[41], and elimination of dI4 presynaptic inhibitory interneurons uncovers forelimb oscillation during reaching[42], phenotypes distinct from that seen after V2a interneuron ablation. Classical studies in cat and primate also speak to the issue of conservation in PN function and behavioural modularity[16,32]. As with V2a neuronal ablation, severing the premotor output of cat PNs undermines reaching but not grasping[25]. The consistency of behavioural defects across mammalian species and interventional approaches suggests that PNs represent a major subpopulation of V2a interneurons involved in the control of reaching. In monkey, however, blocking the output of PNs disrupts both reach and grasp[26], which may reflect an evolutionary elaboration in the wiring of circuits for distal forelimb dexterity[16,43].

Our findings also provide insight into the nature of feedback provided by internal copy pathways. The PN-mediated recruitment of LRN neurons activates motor neurons through a fast cerebellar feedback loop. The most likely substrate for such rapid feedback is the collateral projection of LRN axons to deep cerebellar nuclei, which then recruit reticulospinal neurons that activate motor neurons as well as PNs themselves (Supplementary Discussion)[31,37,44,45]. Yet our findings pose the problem of the logical underpinnings of a circuit in which enhanced PN activity triggers a reinforcing feedback loop that further excites PNs and motor neurons, potentially destabilizing motor output. One potential resolution lies in the fact that we manipulated solely an excitatory subpopulation of PNs—the physiological operation of the PN system may depend critically on parallel recruitment of inhibitory PNs[24,29,46]. Our focus on the most rapid PN response pathway does not preclude the engagement of cerebellar granule and Purkinje circuits that display plasticity during motor adaptation and learning[10,47,48], permitting an additional level of dynamic response to PN activity.

Finally, we consider the merits of an internal copy conveyed from the spinal cord. By virtue of their privileged status as last-order interneurons, and their role as mediators of convergent descending and segmental sensory input, PNs may be afforded access to a degree of integrated premotor information not available upstream. We note that the PN internal copy strategy differs in design from more conventional spinal efference copy systems that lack a premotor output arm, and as a consequence are required to receive and transmit a facsimile of the diverse inputs that impinge on motor neurons[49,50]. Admittedly, PNs provide only one of many convergent inputs to motor neurons[32]. Yet their bifurcating output reduces the burden inherent in conveying an accurate copy of premotor information, simplifying the task of matching premotor signal and internal report.

METHODS SUMMARY

Procedures performed in this study were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee of Columbia University and the ethical committee of Umeå University. See Full Methods for further details.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 August 2013; accepted 13 January 2014.
Published online 2 February 2014.

1. Iwaniuk, A. N. & Whishaw, I. Q. On the origin of skilled forelimb movements. *Trends Neurosci.* **23,** 372–376 (2000).
2. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* **47,** 381–391 (1954).
3. Morasso, P. Spatial control of arm movements. *Exp. Brain Res.* **42,** 223–227 (1981).
4. Lawrence, D. G. & Kuypers, H. G. The functional organization of the motor system in the monkey. I. The effects of bilateral pyramidal lesions. *Brain* **91,** 1–14 (1968).
5. Lawrence, D. G. & Kuypers, H. G. The functional organization of the motor system in the monkey. II. The effects of lesions of the descending brain-stem pathways. *Brain* **91,** 15–36 (1968).
6. Lemon, R. N. Descending pathways in motor control. *Annu. Rev. Neurosci.* **31,** 195–218 (2008).
7. Prablanc, C. & Martin, O. Automatic control during hand reaching at undetected two-dimensional target displacements. *J. Neurophysiol.* **67,** 455–469 (1992).
8. Messier, J. & Kalaska, J. F. Comparison of variability of initial kinematics and endpoints of reaching movements. *Exp. Brain Res.* **125,** 139–152 (1999).
9. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5,** 1226–1235 (2002).
10. Ito, M. *The Cerebellum and Neural Control* (Raven, 1984).
11. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2,** 338–347 (1998).
12. Kawato, M. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9,** 718–727 (1999).
13. Scott, S. H. Optimal feedback control and the neural basis of volitional motor control. *Nature Rev. Neurosci.* **5,** 532–546 (2004).
14. Shadmehr, R. & Krakauer, J. W. A computational neuroanatomy for motor control. *Exp. Brain Res.* **185,** 359–381 (2008).
15. Alstermark, B. & Lundberg, A. in *Muscle Afferents and Spinal Control of Movement* (eds Jami, L., Pierrot-Deseilligny, E. & Zytnicki, D.) 327–354 (Pergamon, 1992).
16. Alstermark, B. & Isa, T. Circuits for skilled reaching and grasping. *Annu. Rev. Neurosci.* **35,** 559–578 (2012).
17. Illert, M., Jankowska, E., Lundberg, A. & Odutola, A. Integration in descending motor pathways controlling the forelimb in the cat. 7. Effects from the reticular formation on C3–C4 propriospinal neurones. *Exp. Brain Res.* **42,** 269–281 (1981).
18. Illert, M., Lundberg, A. & Tanaka, R. Integration in descending motor pathways controlling the forelimb in the cat. 3. Convergence on propriospinal neurones transmitting disynaptic excitation from the corticospinal tract and other descending tracts. *Exp. Brain Res.* **29,** 323–346 (1977).
19. Alstermark, B., Kummel, H., Pinter, M. J. & Tantisira, B. Integration in descending motor pathways controlling the forelimb in the cat. 17. Axonal projection and termination of C3–C4 propriospinal neurones in the C6-Th1 segments. *Exp. Brain Res.* **81,** 447–461 (1990).
20. Alstermark, B., Lindstrom, S., Lundberg, A. & Sybirska, E. Integration in descending motor pathways controlling the forelimb in the cat. 8. Ascending projection to the lateral reticular nucleus from C3–C4 propriospinal also projecting to forelimb motoneurones. *Exp. Brain Res.* **42,** 282–298 (1981).
21. Brodal, A. Spinal afferents to the lateral reticular nucleus of the medulla oblongata in the cat; an experimental study. *J. Comp. Neurol.* **91,** 259–295 (1949).
22. Corvaja, N., Grofova, I., Pompeiano, O. & Walberg, F. The lateral reticular nucleus in the cat—I. An experimental anatomical study of its spinal and supraspinal afferent connections. *Neuroscience* **2,** 537–553 (1977).
23. Arshavsky, Y. I., Gelfand, I. M., Orlovsky, G. N. & Pavlova, G. A. Messages conveyed by spinocerebellar pathways during scratching in the cat. I. Activity of neurons of the lateral reticular nucleus. *Brain Res.* **151,** 479–491 (1978).
24. Alstermark, B. & Ekerot, C. F. The lateral reticular nucleus: a precerebellar center providing the cerebellum with overview and integration of motor functions at systems level. A new hypothesis. *J. Physiol.* **591,** 5453–5458 (2013).
25. Alstermark, B., Lundberg, A., Norrsell, U. & Sybirska, E. Integration in descending motor pathways controlling the forelimb in the cat. 9. Differential behavioural defects after spinal cord lesions interrupting defined pathways from higher centres to motoneurones. *Exp. Brain Res.* **42,** 299–318 (1981).
26. Kinoshita, M. *et al.* Genetic dissection of the circuit for hand dexterity in primates. *Nature* **487,** 235–238 (2012).

©2014 Macmillan Publishers Limited. All rights reserved

27. Al-Mosawie, A., Wilson, J. M. & Brownstone, R. M. Heterogeneity of V2-derived interneurons in the adult mouse spinal cord. *Eur. J. Neurosci.* **26,** 3003–3015 (2007).
28. Lundfald, L. *et al.* Phenotype of V2-derived interneurons and their relationship to the axon guidance molecule EphA4 in the developing mouse spinal cord. *Eur. J. Neurosci.* **26,** 2989–3002 (2007).
29. Pivetta, C., Esposito, M. S., Sigrist, M. & Arber, S. Motor circuit communication matrix from spinal cord to brainstem neurons revealed by developmental origin. *Cell* http://dx.doi.org/10.1016/j.cell.2013.12.014 (2014).
30. Goulding, M. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Rev. Neurosci.* **10,** 507–518 (2009).
31. Alstermark, B. & Ogawa, J. *In vivo* recordings of bulbospinal excitation in adult mouse forelimb motoneurons. *J. Neurophysiol.* **92,** 1958–1962 (2004).
32. Alstermark, B. & Kummel, H. Transneuronal transport of wheat germ agglutinin conjugated horseradish peroxidase into last order spinal interneurones projecting to acromio- and spinodeltoideus motoneurones in the cat. 2. Differential labelling of interneurones depending on movement type. *Exp. Brain Res.* **80,** 96–103 (1990).
33. Whishaw, I. Q. An endpoint, descriptive, and kinematic comparison of skilled reaching in mice (*Mus musculus*) with rats (*Rattus norvegicus*). *Behav. Brain. Res.* **78,** 101–111 (1996).
34. Farr, T. D. & Whishaw, I. Q. Quantitative and qualitative impairments in skilled reaching in the mouse (*Mus musculus*) after a focal motor cortex stroke. *Stroke* **33,** 1869–1875 (2002).
35. Buch, T. *et al.* A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nature Methods* **2,** 419–426 (2005).
36. Farr, T. D., Liu, L., Colwell, K. L., Whishaw, I. Q. & Metz, G. A. Bilateral alteration in stepping pattern after unilateral motor cortex injury: a new test strategy for analysis of skilled limb movements in neurological mouse models. *J. Neurosci. Methods* **153,** 104–113 (2006).
37. Wu, H. S., Sugihara, I. & Shinoda, Y. Projection patterns of single mossy fibers originating from the lateral reticular nucleus in the rat cerebellar cortex and nuclei. *J. Comp. Neurol.* **411,** 97–118 (1999).
38. Zhan, X. & Ryugo, D. K. Projections of the lateral reticular nucleus to the cochlear nucleus in rats. *J. Comp. Neurol.* **504,** 583–598 (2007).
39. Rose, M. F. *et al. Math1* is essential for the development of hindbrain neurons critical for perinatal breathing. *Neuron* **64,** 341–354 (2009).
40. Bizzi, E., Cheung, V. C., d'Avella, A., Saltiel, P. & Tresch, M. Combining modules for movement. *Brain Res. Rev.* **57,** 125–133 (2008).
41. Bui, T. V. *et al.* Circuits for grasping: spinal dI3 interneurons mediate cutaneous control of motor behavior. *Neuron* **78,** 191–204 (2013).
42. Fink, A. J. P. *et al. Spinal Presynaptic Inhibition Promotes Smooth Limb Trajectories During Reaching* (Cosyne, 2013).
43. Isa, T., Ohki, Y., Alstermark, B., Pettersson, L. G. & Sasaki, S. Direct and indirect cortico-motoneuronal pathways and control of hand/arm movements. *Physiology* **22,** 145–152 (2007).
44. Carpenter, M. B. & Nova, H. R. Descending division of the brachium conjunctivum in the cat a cerebello-reticular system. *J. Comp. Neurol.* **114,** 295–305 (1960).
45. Schultz, W., Montgomery, E. B. Jr. & Marini, R. Proximal limb movements in response to microstimulation of primate dentate and interpositus nuclei mediated by brain-stem structures. *Brain* **102,** 127–146 (1979).
46. Alstermark, B., Lundberg, A. & Sasaki, S. Integration in descending motor pathways controlling the forelimb in the cat. 10. Inhibitory pathways to forelimb motoneurones via C3–C4 propriospinal neurones. *Exp. Brain Res.* **56,** 279–292 (1984).
47. Fujita, M. Adaptive filter model of the cerebellum. *Biol. Cybern.* **45,** 195–206 (1982).
48. Dean, P., Porrill, J., Ekerot, C. F. & Jorntell, H. The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nature Rev. Neurosci.* **11,** 30–43 (2010).
49. Lundberg, A. Function of the ventral spinocerebellar tract. A new hypothesis. *Exp. Brain Res.* **12,** 317–330 (1971).
50. Arshavsky, Y. I., Gelfand, I. M., Orlovsky, G. N. & Pavlova, G. A. Messages conveyed by spinocerebellar pathways during scratching in the cat. II. Activity of neurons of the ventral spinocerebellar tract. *Brain Res.* **151,** 493–506 (1978).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Zagoraiou, S. Crone and K. Sharma for the *Chx10-cre* mouse; T. Buch for the *DTR-GFP* construct; K. Deisseroth for the *AAV-FLEX-hChR2-YFP* plasmid; and M. Churchland for providing a low-pass filter. We are grateful to J. Krakauer for discussion of concepts in mammalian motor control and debate about terminology and text, and T. Isa for comments and continuing discussion on primate forelimb movement. R. Axel, M. Churchland, D. Jabaudon, A. Karpova, A. Miri, N. Sawtell and C. Zuker also provided discussion and comments on the manuscript. We are grateful to K. Miao for technical assistance; S. Fageiry for work on retrograde CTB labeling; T. Akay for assistance with the horizontal ladder task; D. Ng for the Cre-expressing plasmid; M. Mendelsohn, N. Zabello and S. Patruni for animal care; E. Danielsson and P. Utsi for engineering support; and B. Han, K. MacArthur, S. Morton and I. Schieren for technical assistance. E.A. is a Howard Hughes Medical Institute Fellow of the Helen Hay Whitney Foundation; J.J. was supported by grants to B.A. from Umeå University; B.A. was supported by grants from the Swedish Research Council; T.M.J. was supported by National Institutes of Health grant NS033245, the Harold and Leila Y. Mathers Foundation and Project A.L.S., and is an investigator of the Howard Hughes Medical Institute.

Author Contributions E.A., B.A. and T.M.J. designed the experiments and analysed data. E.A. performed molecular, anatomical and behavioural experiments. B.A., J.J. and E.A. performed and analysed electrophysiological experiments. E.A., B.A. and T.M.J. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.J. (tmj1@columbia.edu), B.A. (bror.alstermark@umu.se) or E.A. (ea2471@columbia.edu).

©2014 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature13175

Isotopic links between atmospheric chemistry and the deep sulphur cycle on Mars

Heather B. Franz[1,2], Sang-Tae Kim[3], James Farquhar[2], James M. D. Day[4], Rita C. Economos[5], Kevin D. McKeegan[5], Axel K. Schmitt[5], Anthony J. Irving[6], Joost Hoek[2] & James Dottin III[2]

The geochemistry of Martian meteorites provides a wealth of information about the solid planet and the surface and atmospheric processes that occurred on Mars. The degree to which Martian magmas may have assimilated crustal material, thus altering the geochemical signatures acquired from their mantle sources, is unclear[1]. This issue features prominently in efforts to understand whether the source of light rare-earth elements in enriched shergottites lies in crustal material incorporated into melts[1,2] or in mixing between enriched and depleted mantle reservoirs[3]. Sulphur isotope systematics offer insight into some aspects of crustal assimilation. The presence of igneous sulphides in Martian meteorites with sulphur isotope signatures indicative of mass-independent fractionation suggests the assimilation of sulphur both during passage of magmas through the crust of Mars and at sites of emplacement. Here we report isotopic analyses of 40 Martian meteorites that represent more than half of the distinct known Martian meteorites, including 30 shergottites (28 plus 2 pairs, where pairs are separate fragments of a single meteorite), 8 nakhlites (5 plus 3 pairs), Allan Hills 84001 and Chassigny. Our data provide strong evidence that assimilation of sulphur into Martian magmas was a common occurrence throughout much of the planet's history. The signature of mass-independent fractionation observed also indicates that the atmospheric imprint of photochemical processing preserved in Martian meteoritic sulphide and sulphate is distinct from that observed in terrestrial analogues, suggesting fundamental differences between the dominant sulphur chemistry in the atmosphere of Mars and that in the atmosphere of Earth[4].

Sulphur isotope analyses were undertaken using sequential chemical extraction from whole-rock meteorite samples followed by dual-inlet isotope ratio mass spectrometry and *in situ* secondary ion mass spectrometry (SIMS). The abundance of sulphur extracted from each fraction is reported in Supplementary Table 1 and illustrated in Extended Data Fig. 1. Isotope data are reported in per mil using δ and Δ notation (Methods and Supplementary Table 2a) and were normalized to measurements of Cañon Diablo Troilite (CDT). A subset of the data, identified with anomalous sulphur isotopic signatures, is reported in Supplementary Table 2b. Anomalous sulphur isotope effects, which produce mass-independent fractionation (MIF), are those in which isotope selection depends on factors in addition to the classical (mass-dependent) selection rules associated with molecular vibrations, rotations and translations. These effects are typically characterized by significant deviations of $\Delta^{33}S$, $\Delta^{36}S$ or both, but we note that small variations in $\Delta^{33}S$ and $\Delta^{36}S$ may be produced by mixing or distillation processes involving large (tens of per mil) mass-dependent variation in $\delta^{34}S$ (Supplementary Information and Extended Data Fig. 3). The magnitude of $\Delta^{33}S$ and the absence of significant $\delta^{34}S$ variation in Martian meteorites rule out such mixing and distillation and have led to the suggestion that the Martian signatures reflect a mass-independent effect, attributed to ultraviolet photochemistry[5,6]. We discuss below how the MIF signatures observed in this study provide insight into processes of assimilation of sulphur into magma during transport (observed in shergottites), processes of assimilation of sulphur salts or sulphur-bearing fluids at the time the flows were emplaced (observed in nakhlites), and in some cases processes that introduce MIF sulphur signatures through secondary alteration processes on Mars.

The shergottites are igneous rocks that lack evidence of extensive secondary mineralization[7]. We extracted acid-volatile sulphur (AVS, predominantly monosulphide) from shergottites representing each mineralogical and trace element category identified in the literature. The isotopic composition of the AVS fraction reveals uniformity among these samples (Fig. 1) and establishes a baseline for interpretation of Martian sulphur isotopic data. The weighted mean of analyses yields -0.17 ± 0.11‰ for $\delta^{34}S$, 0.009 ± 0.001‰ for $\Delta^{33}S$ and 0.02 ± 0.19‰ for $\Delta^{36}S$ (see Methods for details; uncertainties represent 2 s.e.m.). The data do not follow a normal distribution and yield a standard deviation of 0.013‰ for $\Delta^{33}S$, which is ~1.5 times that predicted from analytical uncertainty alone. The variability in $\Delta^{33}S$ for the sample population suggests augmentation by sulphur with anomalous $\Delta^{33}S$ that did not originate in the mantle. However, the $\Delta^{33}S$ of the mean is tightly constrained and close to the value measured for CDT and bulk chondritic sulphur (average, 0.002 ± 0.005‰ (s.e.m.) ± 0.025‰ (s.d.); refs 8, 9). Data for groups of enriched versus depleted shergottites and mafic versus ultramafic shergottites are distributed about the mean, and the sulphur isotopic composition of Y980459, which is one of the most primitive shergottites and is thus inferred to represent Martian mantle composition[10], notably lies close to the mean. Although other processes may have influenced the sulphur isotopic composition of shergottites, the mean value of 0.009‰ provides a reasonable measure of the juvenile $\Delta^{33}S$ of Mars. This signal is less ^{33}S-enriched than those observed in some achondrite groups and minor components of chondrites that have been interpreted to reflect heterogeneity in planetary precursor materials[9,11,12]. The variability in $\Delta^{33}S$ of Martian sulphur, chondritic sulphur[8,9] and terrestrial mantle sulphur[13] is less than 0.01‰, implying that sulphur was well mixed in the most abundant inner Solar System materials. The $\delta^{34}S$ values derived from our shergottite AVS data are also homogeneous and yield a mean value within error of the CDT value, suggesting efficient mixing in the Martian mantle and little or no fractionation of sulphur isotopes during core formation.

Three shergottites have resolvable mass-independent AVS $\Delta^{33}S$ compositions, shown with open symbols in Fig. 1a. These are NWA 2990, NWA 5960 (paired with the former) and Los Angeles, with $\Delta^{33}S$ values of 0.063 ± 0.014‰, 0.093 ± 0.008‰ and 0.054 ± 0.008‰, respectively. Because MIF sulphur signatures are not produced by magmatic processes, these observations suggest that sulphur was fractionated by ultraviolet photochemistry in the Martian atmosphere[5,6], deposited onto the surface and ultimately assimilated into melts, where it was reduced and incorporated into igneous sulphides. NWA 2990, NWA 5960 and Los Angeles also exhibit trace element and radiogenic isotopic characteristics

[1]Center for Research and Exploration in Space Science and Technology, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [2]Department of Geology and Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742, USA. [3]School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [4]Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA. [5]Department of Earth and Space Sciences, University of California, Los Angeles, California 90095, USA. [6]Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA.

©2014 Macmillan Publishers Limited. All rights reserved

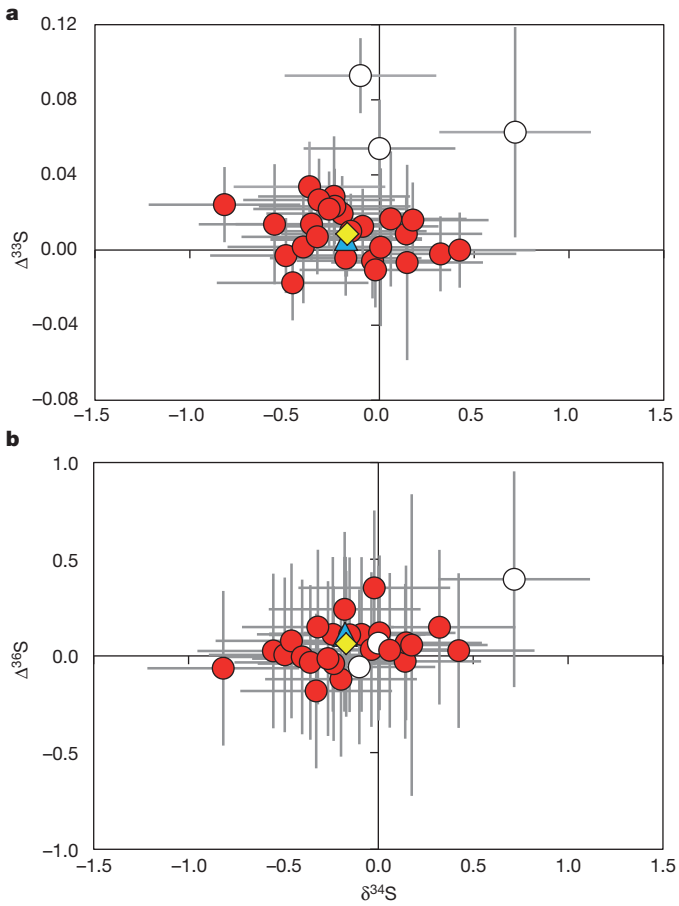


Figure 1 | Sulphur isotopic compositions of shergottite AVS. **a**, $\Delta^{33}S$ versus $\delta^{34}S$; **b**, $\Delta^{36}S$ versus $\delta^{34}S$. Yellow diamond, weighted mean composition; blue triangle, Y980459; filled circles, samples included in the weighted mean; open circles, samples not included in the weighted mean due to anomalous signatures. Error bars, 2 s.d.

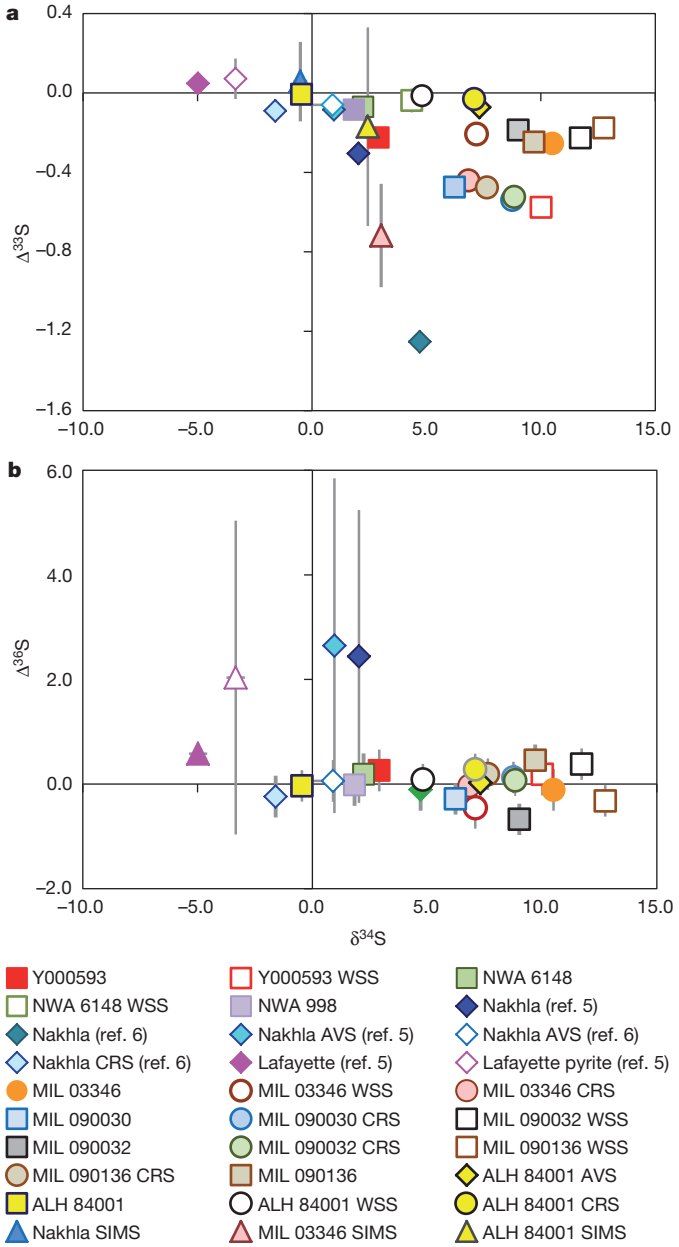


Figure 2 | Sulphur isotopic compositions of nakhlites and ALH 84001. **a**, $\Delta^{33}S$ versus $\delta^{34}S$; **b**, $\Delta^{36}S$ versus $\delta^{34}S$. AVS, acid-volatile sulphur; CRS, chromium-reducible sulphur; WSS, water-soluble sulphate; otherwise (no suffix in key), acid-soluble sulphate. Error bars, 2 s.d. Results of previous studies[5,6] are shown for comparison. **a** also includes weighted mean values for SIMS analyses of Nakhla, MIL 03346 and ALH 84001.

that some workers have interpreted to indicate incorporation of crustal material[14–17]. The coincidence of anomalous sulphide with other geochemical characteristics, including evidence for crystallization under relatively oxidizing conditions[14–17], suggests a link between anomalous sulphur in igneous minerals and assimilation of crustal material into the melts.

Chemical extractions of sulphide (AVS, chromium-reducible sulphur or both) from Nakhla[5,6] ($\Delta^{33}S = -0.08$ ‰), Allan Hills (ALH) 84001 ($\Delta^{33}S = -0.072 \pm 0.008$‰), and Miller Range (MIL) 03346 and three of its pairs ($\Delta^{33}S = -0.434 \pm 0.008$‰ to -0.538 ± 0.008‰) reveal MIF sulphur signatures (Fig. 2), suggesting that crustal material was assimilated by their parent melts. SIMS analyses of Nakhla pyrrhotite yielded a mean $\Delta^{33}S$ value within error of chemical extraction analyses, and SIMS analyses of MIL 03346 pyrrhotite yielded a mean $\Delta^{33}S$ value of -0.72 ± 0.13‰ (Supplementary Table 3), which is slightly more negative than is observed in chemical extractions. Furthermore, analyses of $\Delta^{33}S$ for individual grains of MIL 03346 ranged from -0.30 ± 0.13‰ to -1.11 ± 0.13‰, representing a level of variation outside the range of analytical uncertainty, which we interpret to reflect real differences in isotopic composition.

MIL 03346 is an unusual, fast-cooled nakhlite possessing skeletal iron oxide grains and interspersed sulphide in the mesostasis[18] (Fig. 3 and Extended Data Figs 4–6), offering a rare perspective on the petrogenesis of its anomalous pyrrhotite. The heterogeneous, non-zero $\Delta^{33}S$ of pyrrhotite grains in MIL 03346 suggests a non-equilibrium process associated with assimilation of sulphur salts during emplacement of the magmatic flow from which this meteorite was derived. Assimilation may also be reflected in the highly oxidized state of this meteorite, because reduction of one mole of S(VI) to S(II−) consumes eight moles of electrons and has the potential to oxidize eight moles of Fe(II) to Fe(III). Assimilation of an oxoanion salt such as sulphate, which would lead to iron loss from pyroxene and formation of magnetite and sulphide, is consistent with both the abundance of iron oxide (magnetite) in MIL 03346 and the variable $\Delta^{33}S$ observed in its sulphide minerals. We interpret the differences in $\Delta^{33}S$ observed in pyrrhotites of MIL 03346 and Nakhla to reflect local differences in the proportion and sulphur isotopic composition of assimilated sulphate.

ALH 84001 yielded two pyrite grains with $\Delta^{33}S$ anomalies, one of which was resolvable at the 2 s.d. level. The remaining grains showed mass-dependent isotopic composition, producing a weighted mean $\Delta^{33}S$ of -0.17 ± 0.13‰. As described above, monosulphides of ALH 84001 possess a MIF signature, suggesting assimilation of crustal sulphur into the parent magma. However, pyrite in ALH 84001 is of secondary origin[19,20]. Variations in $\delta^{34}S$ and $\Delta^{33}S$ observed in pyrite grains by SIMS are consistent with previous reports of sulphur isotopic variability in pyrite from

©2014 Macmillan Publishers Limited. All rights reserved

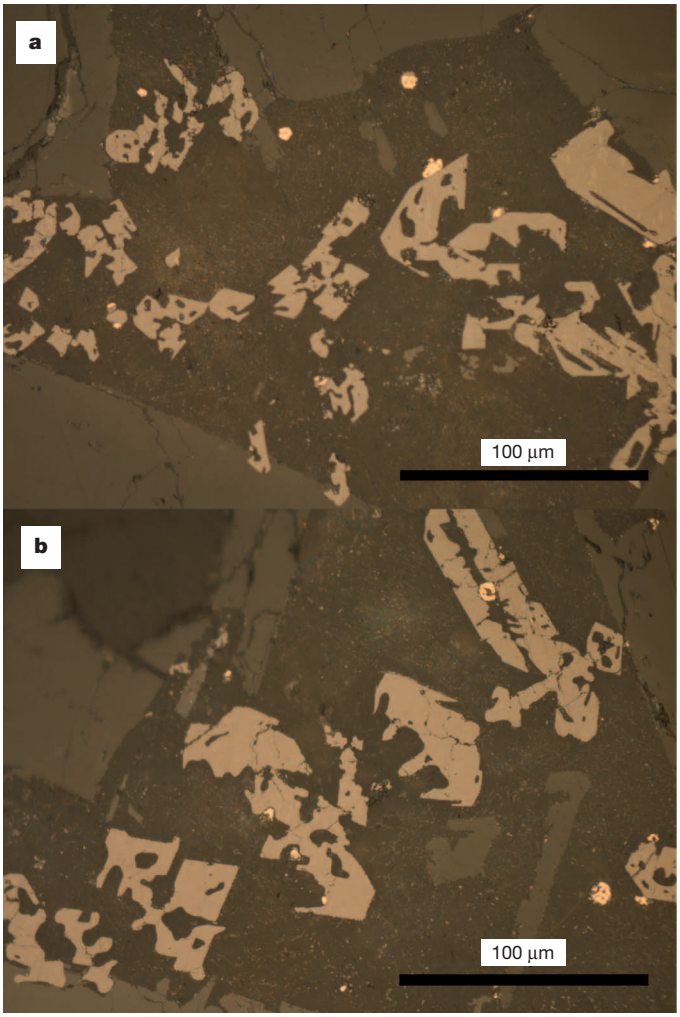


Figure 3 | Reflected-light images of MIL 090030, paired with MIL 03346. The images illustrate the association between pyrrhotite (bright yellow) and skeletal magnetite (light grey) grains in the intercumulus matrix.

this meteorite[20,21]. There is no evidence that the analysed grains are in brecciated or fragmented zones. Instead, we observed close association of the sulphides with Cr-spinel and maskelynite, as reported in ref. 19. These pyrites have sulphur-poor compositions (Supplementary Table 4), suggesting a possible origin in incomplete sulphurization of igneous pyrrhotite by hydrothermal fluid bearing crustal-derived sulphur—a second mechanism for imparting heterogeneous MIF sulphur isotopic signatures to the meteorites.

Sulphate fractions of the meteorites (Fig. 2) reveal another facet of the Martian sulphur cycle. Anomalous sulphate was detected in six shergottites, Chassigny and all nakhlites analysed. Whereas replicate analyses of NWA 5298 revealed positive $\Delta^{33}S$ (0.146 ± 0.008‰ and 0.170 ± 0.008‰) in sulphate but insufficient AVS for accurate isotope ratio measurements, the $\Delta^{33}S$ signals in sulphate from NWA 2990, NWA 5960 and Los Angeles are within analytical uncertainty of their respective sulphide $\Delta^{33}S$ values, suggesting a genetic link between sulphate and sulphide fractions. This sulphate could reflect incomplete chemical reduction of assimilated crustal sulphate within the magma or an origin as a subsequent sulphide oxidation product. In contrast, sulphates from two other shergottites (NWA 5718 and LAR 06319, with $\Delta^{33}S$ = 0.260 ± 0.008‰ and 0.057 ± 0.008‰, respectively) and Chassigny ($\Delta^{33}S$ = 0.033 ± 0.008‰), which do not have sulphide with anomalous $\Delta^{33}S$, suggest that mass-independent sulphate with positive $\Delta^{33}S$ was introduced through post-crystallization aqueous alteration rather than during the magmatic stage.

Similarly, the MIF signature in nakhlite sulphate suggests a combination of sulphate introduced through post-crystallization aqueous alteration and sulphate assimilated by melts. Analyses of sulphate from Nakhla yield $\Delta^{33}S$ values as negative as −1.25 ± 0.01‰ (ref. 6). Sulphates of MIL 03346 (and pairs), Y000593, NWA 6148 and NWA 998 have sulphur isotopic characteristics consistent with Nakhla, with $\Delta^{33}S$ ranging from −0.038 ± 0.030‰ to −0.576 ± 0.008‰ and $\Delta^{36}S$ near zero. Correlating $\Delta^{33}S$ with the relative positions of the nakhlites in a cumulate pile formed by gravitational settling of crystals within a magmatic flow[18,22–25] shows that meteorites with the largest negative $\Delta^{33}S$ magnitudes (Nakhla, MIL 03346 and Y000593) derive from intermediate depths and that the smallest values of $\Delta^{33}S$ are seen in meteorites from the greatest and shallowest depths (NWA 998 and NWA 6148). The variations in sulphate $\Delta^{33}S$ among the nakhlites imply differences in composition of their alteration fluids.

Different scenarios have been proposed for nakhlite secondary alteration, including alteration by a single event (see, for example, ref. 26) or by multiple events (see, for example, ref. 27). Multiple events could have introduced fluids characterized by different $\Delta^{33}S$ values. Evidence that the Lafayette meteorite experienced alterations unique among the nakhlites, including positive $\Delta^{33}S$ anomalies in its sulphate and pyrite[5], supports the case for multiple events. Although sulphur isotope heterogeneity may be simpler to explain by multiple alteration events, these data do not uniquely support that scenario. Petrological models proposing that the nakhlites were formed by multiple magmatic events instead of a single lava flow (see, for example, refs 23, 28, 29) would assert different constraints on the timing and mode of alteration (Supplementary Information) and are also permitted by the sulphur isotope data.

The data presented here provide insight into the architecture of the Martian sulphur cycle, suggesting the presence of a persistent process that produced geographically distributed sulphur with anomalous $\Delta^{33}S$. Geographical variations in sulphur isotopic composition are supported by different ejection ages of groups of Martian meteorites, indicating that a number of distinct impact events launched meteorites from various locations on the Martian surface (see, for example, ref. 30). The variation in $\Delta^{33}S$ with an absence of significant anomalies in $\Delta^{36}S$ (Fig. 4) implies MIF production by different pathways from those that operated on the early Earth and in Earth's stratosphere today[4,31]. Furthermore, the positive $\Delta^{33}S$ signal in shergottites, Chassigny and Lafayette contrasts the negative $\Delta^{33}S$ observed in the other nakhlites and ALH 84001. This variation in $\Delta^{33}S$ could indicate a shift in composition of the Martian

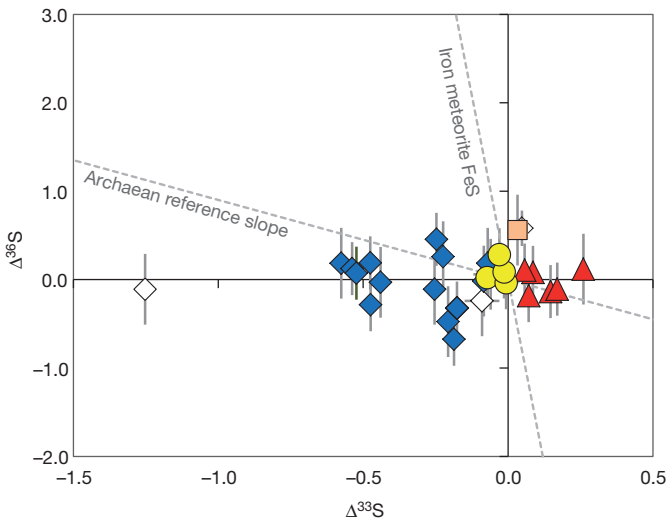


Figure 4 | Covariation between $\Delta^{33}S$ and $\Delta^{36}S$ in different groups of Martian meteorites. Diamonds, nakhlites; circles, ALH 84001; triangles, shergottites; square, Chassigny. Martian meteorite data from other studies are included[5,6] (open symbols). Dashed lines depict arrays formed by Archaean terrestrial samples and iron meteorite FeS for comparison. (See Extended Data Fig. 2 for details.) Error bars, 2 s.d.

©2014 Macmillan Publishers Limited. All rights reserved

atmosphere or in the speciation of sulphur-bearing volcanic gases over time, or it could reflect complementary sulphur products generated by a single photochemical process, as suggested by the uniform array formed by the isotopic signatures of these meteorites in the $\Delta^{36}S-\Delta^{33}S$ space.

The prevalence of sulphur extracted as sulphate implies the predominance of sulphate-formation pathways in the generation or preservation, or both, of MIF sulphur on Mars. Evidence for the transfer of sulphate with anomalous $\Delta^{33}S$ to the regolith is present in meteorites with a variety of protolith ages (ALH 84001, the nakhlites, Chassigny and six shergottites), suggesting that production of mass-independent signatures was sustained throughout much of Mars' history. Assimilation of sulphate with anomalous $\Delta^{33}S$ by Martian melts also seems to have been a common occurrence, as evidenced by isotopic anomalies in sulphide extracted from ALH 84001, several nakhlites and three shergottites (NWA 2990, NWA 5960 and Los Angeles), as well as the small $\Delta^{33}S$ variability in the population of shergottites. These observations imply the cycling of sulphur between the regolith and magma that erupted at, or near, the Martian surface.

Anomalous sulphur transferred from surface materials into Martian igneous minerals provides a sensitive marker for crustal assimilative processes, but the absence of a MIF signal in igneous sulphides does not constitute evidence for the absence of assimilation. For example, the composition of light rare-earth elements in early-trapped melt inclusions in LAR 06319 is similar to that in whole rock, indicating that this meteorite probably derived its enrichment properties from partial melting of an enriched and oxidized mantle reservoir, which did not involve assimilation of crustal material[32]. Although this interpretation is in accord with our evidence that MIF sulphur was added to LAR 06319 during post-crystallization alteration, the sulphur isotope ratios cannot rule out assimilation of crustal sulphur that did not bear a MIF signature. Considered in conjunction with data for other isotopic systems and trace element characteristics, sulphur isotope systematics present a powerful tool for reconstructing the geological history of Mars.

METHODS SUMMARY

A sequential chemical extraction procedure was used to obtain sulphur isotope ratios for different mineral phases in powdered whole-rock samples. Each sample powder was pre-treated by sonicating in Milli-Q water to extract water-soluble sulphate, which was converted to barium sulphate and dried for later reduction. The sample powder was then processed with a series of three acidic solutions to extract different sulphur-bearing phases. In the first step, the sample was heated with 5 N HCl under flowing N_2 gas. AVS in the sample, presumed to consist primarily of monosulphides, reacted with the HCl to produce H_2S gas, which was collected in an acidic capture solution containing $AgNO_3$. In two subsequent steps, additional reduction solutions were injected into the boiling flask for sequential extraction of acid-soluble sulphate and chromium-reducible sulphur from the sample.

The H_2S produced in each step of the extraction process reacted with $AgNO_3$ in the capture solution to form Ag_2S, which was rinsed with Milli-Q water and a 1 M NH_4OH solution, and then dried. Samples of Ag_2S were reacted with ~10-fold stoichiometric excess of pure F_2 at ~250 °C. Product SF_6 was purified by both cryogenic and gas chromatographic techniques, and sulphur isotope abundances were measured by monitoring mass/charge ratios m/z 127, 128, 129 and 131 ($^{32}SF_5^+$, $^{33}SF_5^+$, $^{34}SF_5^+$ and $^{36}SF_5^+$, respectively) with a ThermoFinnigan MAT 253 mass spectrometer. Uncertainties were estimated from repeated measurements of International Atomic Energy Agency reference standards and CDT, and are generally less than 0.15‰, 0.008‰ and 0.15‰ (1 s.d.) for $\delta^{34}S$, $\Delta^{33}S$ and $\delta^{36}S$, respectively. Data were normalized to measurements of CDT.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 September 2013; accepted 14 February 2014.

1. Jones, J. H. Isotopic relationships among the shergottites, the nakhlites, and Chassigny. *Proc. Lunar Planet. Sci. Conf.* **19,** 465–474 (1989).
2. Herd, C. D. K., Borg, L. E., Jones, J. H. & Papike, J. J. Oxygen fugacity and geochemical variations in the Martian basalts: implications for Martian basalt petrogenesis and the oxidation state of the upper mantle of Mars. *Geochim. Cosmochim. Acta* **66,** 2025–2036 (2002).
3. Borg, L. E., Nyquist, L. E., Taylor, L. A., Wiesmann, H. & Shih, C.-Y. Constraints on Martian differentiation processes from Rb-Sr and Sm-Nd isotopic analyses of the basaltic shergottite QUE 94201. *Geochim. Cosmochim. Acta* **61,** 4915–4931 (1997).
4. Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289,** 756–758 (2000).
5. Farquhar, J., Savarino, J., Jackson, T. L. & Thiemens, M. H. Evidence of atmospheric sulphur in the Martian regolith from sulphur isotopes in meteorites. *Nature* **404,** 50–52 (2000).
6. Farquhar, J., Kim, S.-T. & Masterson, A. Implications from sulfur isotopes of the Nakhla meteorite for the origin of sulfate on Mars. *Earth Planet. Sci. Lett.* **264,** 1–8 (2007).
7. McSween, H. Y. What we have learned about Mars from SNC meteorites. *Meteoritics* **29,** 757–779 (1994).
8. Gao, X. & Thiemens, M. H. Isotopic composition and concentration of sulfur in carbonaceous chondrites. *Geochim. Cosmochim. Acta* **57,** 3159–3169 (1993).
9. Rai, V. K. & Thiemens, M. H. Mass independently fractionated sulfur components in chondrites. *Geochim. Cosmochim. Acta* **71,** 1341–1354 (2007).
10. Mikouchi, T. *et al.* Yamato 980459: mineralogy and petrology of a new shergottite-related rock from Antarctica. *Antarct. Meteorite Res.* **17,** 13–34 (2004).
11. Gao, X. & Thiemens, M. H. Systematic study of sulfur isotopic composition in iron meteorites and the occurrence of excess ^{33}S and ^{36}S. *Geochim. Cosmochim. Acta* **55,** 2671–2679 (1991).
12. Rai, V. K., Jackson, T. L. & Thiemens, M. H. Photochemical mass-independent sulfur isotopes in achondritic meteorites. *Science* **309,** 1062–1065 (2005).
13. Labidi, J., Cartigny, P. & Moreira, M. Non-chondritic sulphur isotope composition of the terrestrial mantle. *Nature* **501,** 208–211 (2013).
14. Gross, J. *et al.* Petrography, mineral chemistry, and crystallization history of olivine-phyric shergottite NWA 6234: a new melt composition. *Meteorit. Planet. Sci.* **48,** 854–871 (2013).
15. Irving, A. J., Herd, C. D. K., Gellissen, M., Kuehner, S. M. & Bunch, T. E. Paired fine-grained, permafic olivine-phyric shergottites Northwest Africa 2990/5960/6234/6710: trace element evidence for a new type of Martian mantle source or complex lithospheric assimilation process. *Meteorit. Planet. Sci.* **46,** A108 (2011).
16. Rubin, A. E. *et al.* Los Angeles: the most differentiated basaltic Martian meteorite. *Geology* **28,** 1011 (2000).
17. Warren, P. H., Greenwood, J. P. & Rubin, A. E. Los Angeles: a tale of two stones. *Meteorit. Planet. Sci.* **39,** 137–156 (2004).
18. Day, J. M. D., Taylor, L. A., Floss, C. & McSween, H. Y. Jr. Petrology and chemistry of MIL 03346 and its significance in understanding the petrogenesis of nakhlites on Mars. *Meteorit. Planet. Sci.* **41,** 581–606 (2006).
19. Mittlefehldt, D. W. ALH 84001, a cumulate orthopyrocenite member of the Martian meteorite clan. *Meteoritics* **29,** 214–221 (1994).
20. Greenwood, J. P., Mojzsis, S. J. & Coath, C. D. Sulfur isotopic compositions of individual sulfides in Martian meteorites ALH 84001 and Nakhla: implications for crust-regolith exchange on Mars. *Earth Planet. Sci. Lett.* **184,** 23–35 (2000).
21. Shearer, C. K., Layne, G. D., Papike, J. J. & Spilde, M. N. Sulfur isotopic systematics in alteration assemblages in Martian meteorite Allan Hills 84001. *Geochim. Cosmochim. Acta* **60,** 2921–2926 (1996).
22. Mikouchi, T., Miyamoto, M., Koizumi, E., Makishima, J. & McKay, G. Relative burial depths of nakhlites: an update. *Proc. Lunar Planet. Sci. Conf.* **37,** abstr. 1865 (2006).
23. Jambon, A. *et al.* Northwest Africa 5790. Top sequence of the nakhlite pile. *Proc. Lunar Planet. Sci. Conf.* **41,** 1696 (2010).
24. Mikouchi, T. & Miyamoto, M. Comparative cooling rates of nakhlites as inferred from iron-magnesium and calcium zoning of olivines. *Proc. Lunar Planet. Sci. Conf.* **31,** abstr. 1343 (2002).
25. Udry, A., McSween, H. Y., Lecumberri-Sanchez, P. & Bodnar, R. J. Paired nakhlites MIL 090030, 090136, and 03346: insights into the Miller Range parent meteorite. *Meteorit. Planet. Sci.* **47,** 1575–1589 (2012).
26. Bridges, J. C. & Grady, M. M. Evaporite mineral assemblages in the nakhlite (Martian) meteorites. *Earth Planet. Sci. Lett.* **176,** 267–279 (2000).
27. Hallis, L. J. & Taylor, G. J. Comparisons of the four Miller Range nakhlites, MIL 03346, 090030, 090032 and 090136: textural and compositional observations of primary and secondary mineral assemblages. *Meteorit. Planet. Sci.* **46,** 1787–1803 (2011).
28. Shirai, N. & Ebihara, M. Chemical characteristics of nakhlites: implications to the geological setting for nakhlites. *Proc. Lunar Planet. Sci. Conf.* **39,** 1643 (2008).
29. Shih, C.-Y., Nyquist, L. E., Reese, Y. & Jambon, A. Sm-Nd isotopic studies of two nakhlites, NWA 5790 and Nakhla. *Proc. Lunar Planet. Sci. Conf.* **41,** 1367 (2010).
30. McSween, H. Y. in *The Martian Surface: Composition, Mineralogy, and Physical Properties* (ed. Bell, J. F.) 383–396 (Cambridge Univ. Press, 2008).
31. Baroni, M., Savarino, J., Cole-Dai, J., Rai, V. K. & Thiemens, M. H. Anomalous sulfur isotope compositions of volcanic sulfate over the last millennium in Antarctic ice cores. *J. Geophys. Res.* **113,** D20112 (2008).
32. Basu Sarbadhikari, A., Day, J. M. D., Liu, Y., Rumble, D. & Taylor, L. A. Petrogenesis of olivine-phyric shergottite Larkman Nunatak 06319: implications for enriched components in Martian basalts. *Geochim. Cosmochim. Acta* **73,** 2190–2214 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the Meteorite Working Group, L. Welzenbach, T. McCoy, S. Ralew, M. N. Rao, L. Nyquist, J. Zipfel, C. Smith, H. Kojima, A. Treiman,

©2014 Macmillan Publishers Limited. All rights reserved

T. Bunch and B. Zanda for providing meteorite samples analysed in this study. We also thank P. Piccoli for assistance with electron microprobe analyses. The manuscript benefited from independent reviews by M. Thiemens, P. Cartigny, S. Ono, D. Johnston and B. Wing during revision. The UCLA ion microprobe facility is partly supported by a grant from the US National Science Foundation Instrumentation and Facilities Program. This work was supported by NASA Cosmochemistry grants NNX09AF72G and NNX13AL13G to J.F.

Author Contributions H.B.F. and J.F. designed the study and wrote the manuscript. All authors contributed to data collection, data interpretation and editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.B.F. (heather.b.franz@nasa.gov).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13017

Coherent suppression of electromagnetic dissipation due to superconducting quasiparticles

Ioan M. Pop[1]*, Kurtis Geerlings[1]*, Gianluigi Catelani[1,2], Robert J. Schoelkopf[1], Leonid I. Glazman[1] & Michel H. Devoret[1]

Owing to the low-loss propagation of electromagnetic signals in superconductors, Josephson junctions constitute ideal building blocks for quantum memories, amplifiers, detectors and high-speed processing units, operating over a wide band of microwave frequencies. Nevertheless, although transport in superconducting wires is perfectly lossless for direct current, transport of radio-frequency signals can be dissipative in the presence of quasiparticle excitations above the superconducting gap[1]. Moreover, the exact mechanism of this dissipation in Josephson junctions has never been fully resolved experimentally. In particular, Josephson's key theoretical prediction that quasiparticle dissipation should vanish in transport through a junction when the phase difference across the junction is π (ref. 2) has never been observed[3]. This subtle effect can be understood as resulting from the destructive interference of two separate dissipative channels involving electron-like and hole-like quasiparticles. Here we report the experimental observation of this quantum coherent suppression of quasiparticle dissipation across a Josephson junction. As the average phase bias across the junction is swept through π, we measure an increase of more than one order of magnitude in the energy relaxation time of a superconducting artificial atom. This striking suppression of dissipation, despite the presence of lossy quasiparticle excitations above the superconducting gap, provides a powerful tool for minimizing decoherence in quantum electronic systems and could be directly exploited in quantum information experiments with superconducting quantum bits.

Despite the success of Josephson's theoretical predictions, the tunnelling current across a Josephson junction in the presence of quasiparticles is still a subject of controversy[3]. It is customary to distinguish four contributions to the current through a superconducting tunnel junction, associated with Cooper pairs and quasiparticles[2,3]: (1) the current of Cooper pairs, (2) the dispersive quasiparticle term, (3) the phase-independent dissipative term, and (4) the dissipative term proportional to the cosine of the phase difference. Contributions (1) and (2) comprise the supercurrent (that is, the Josephson current), which was measured shortly after its prediction and is nowadays extensively used in applications[4]. The cosine term, also known as quasiparticle-pair interference[5], has eluded conclusive experimental observation[3]. Only very recently, thermal transport measurements showed the effect of quasiparticle-pair interference on the heat conductivity of a superconducting quantum interference device[6]. In a phase-biased Josephson junction, we expect the cosine and dissipative quasiparticle terms in the tunnelling current to add coherently and display a periodic dependence on flux[7–9].

The physical significance of the contributions (1)–(4) to the tunnelling current is best illustrated by the linear response of the junction to an a.c. voltage bias of frequency ω, $V(\omega)$, at a given d.c. phase bias, φ. This junction admittance, relating current and voltage by $I(\omega,\varphi) = Y(\omega,\varphi)V(\omega)$, takes the form

$$Y(\omega,\varphi) = Y_{\mathrm{qp}}(\omega)\frac{1+\varepsilon\cos\varphi}{2} - i\frac{\cos\varphi}{L_{\mathrm{J}}\omega}\left(1-2x_{\mathrm{qp}}^{\mathrm{A}}\right) \qquad (1)$$

The purely imaginary last term in equation (1) comes from the dissipationless supercurrent through a Josephson junction (contributions (1) and (2)), where L_{J} is known as the Josephson inductance. The factor $x_{\mathrm{qp}}^{\mathrm{A}}$ accounts for the occupation of the 'Andreev' bound states, which carry the supercurrent[10,11]; writing contribution (2) in this form allows for a generalization to a non-equilibrium distribution of quasiparticles (Methods). The complex function $Y_{\mathrm{qp}}(\omega)$ depends on the bulk quasiparticle population through the density x_{qp} (normalized by the Cooper pair density)[7], and its real part accounts for dissipation in the junction.

According to standard Bardeen–Cooper–Schrieffer theory, for junctions with identical electrode materials, the ε prefactor in the '$1+\varepsilon\cos\varphi$' term (corresponding to contributions (3) and (4)) approaches unity at temperatures well below the critical temperature for superconductivity[12] and we expect the quasiparticle dissipation to be suppressed at $\varphi = \pi$ (ref. 8). There have been several attempts to measure the sign and amplitude of ε (refs 13, 14); however, owing to the intrinsic dissipative nature of the d.c. experimental methods available in the past, the results were difficult to interpret and proved inconclusive[3]. In this Letter we present direct measurements of dissipation as a function of the phase bias, φ, over a Josephson junction. We observe a sharp decrease in dissipation at $\varphi = \pi$, thus putting stringent bounds on the value of ε.

In principle, for commonly used superconductors such as aluminium and niobium, the quasiparticle fraction, x_{qp}, should ideally reach its thermal equilibrium value, which is practically zero ($x_{\mathrm{qp}}^{\mathrm{th}} \ll 10^{-20}$) at typical dilution refrigerator temperatures of $\lesssim$40 mK. However, in recent years it has become increasingly clear that non-thermal quasiparticles can persist in superconducting circuits even at very low temperatures, inducing relaxation and dephasing[15–20]. Using a fluxonium artificial atom[21], which consists of a Josephson junction shunted by a superinductor[22] (Fig. 1b), we directly measure the dissipation across a flux-biased Josephson junction due to these non-thermal quasiparticles. Superconducting quantum bits (qubits) are ideal testing systems for different dissipation mechanisms. They can be manipulated and measured at the single-photon level and their susceptibility to different loss mechanisms can be tuned *in situ*. We measure the dissipation rate by preparing the qubit in an excited state and observing its energy relaxation. In Fig. 1a, we present the schematic mechanism of quasiparticle dissipation across a phase-biased Josephson junction, the main element of a fluxonium artificial atom, whose first two energy levels constitute a qubit. A quasiparticle from the left electrode can tunnel through the junction while absorbing the qubit energy, resulting in qubit relaxation to the ground state. The quasiparticle can tunnel either as an electron or as a hole, with probability amplitude phases $+\varphi/2$ or $-\varphi/2$, respectively[7]. Because these two events are indistinguishable, the probability amplitudes add coherently to produce the '$1+\cos\varphi$' term in the total probability: $|e^{i\varphi/2} + e^{-i\varphi/2}|^2 = 2(1+\cos\varphi)$.

To identify and quantify the multiple decay mechanisms acting on the qubit, we need to be able to decompose the total relaxation rate into its constituents. The fluxonium qubit (Fig. 1b) is a particularly attractive tool because its susceptibility to each loss mechanism has a distinctive functional dependence on applied magnetic flux, Φ_{ext} (ref. 23),

[1]Department of Applied Physics, Yale University, 15 Prospect Street, New Haven, Connecticut 06511, USA. [2]Forschungszentrum Jülich, Peter Grünberg Institut (PGI-2), 52425 Jülich, Germany.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

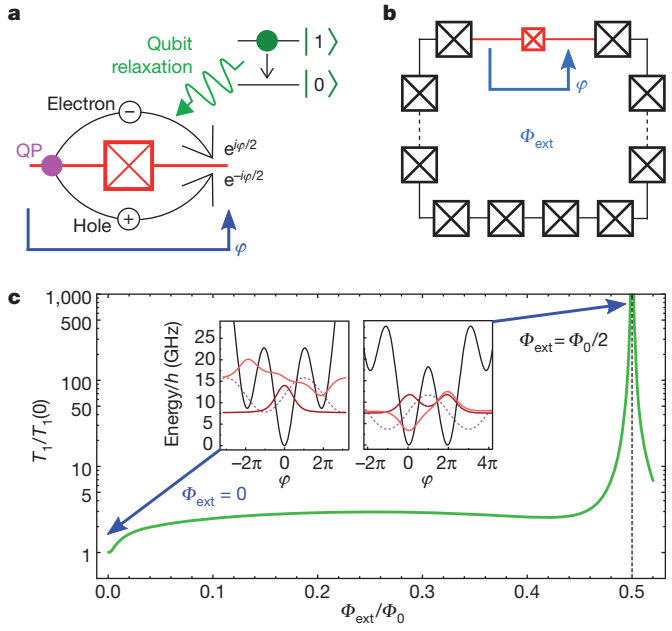


Figure 1 | Phase-dependent quasiparticle dissipation in a Josephson junction. **a**, In the case of a Josephson junction biased at phase φ, quasiparticles (QPs) receiving a quanta of excitation from the qubit or environment can tunnel across the junction either as electrons, acquiring a phase $+\varphi/2$, or as holes, acquiring a phase $-\varphi/2$. The φ dependence of quasiparticle dissipation results from the interference of these two indistinguishable paths. **b**, Schematic representation of the fluxonium qubit: the phase-slip junction (red) is shunted by an array of bigger junctions (black) that form the superinductor. In the fluxonium qubit, the external magnetic flux, Φ_{ext}, sets the average phase difference, φ, across the phase-slip junction. **c**, Flux dependence of the fluxonium relaxation time, T_1, due to quasiparticle loss across the phase-slip junction. Quasiparticle loss is suppressed at $\Phi_{ext}/\Phi_0 = 0.5$, where the average phase difference over the small junction is π. The insets show the potential landscape (black), the qubit ground-state ($|0\rangle$; dark red) and excited-state ($|1\rangle$; light red) wavefunctions, and the quasiparticle dissipation operator[7], $\sin(\varphi/2)$ (dashed magenta). The T_1 divergence at $\Phi_{ext}/\Phi_0 = 0.5$ is explained by the symmetry of $\sin(\varphi/2)$ around $\varphi = \pi$, together with the respective even and odd parities of $|0\rangle$ and $|1\rangle$, which lead to the vanishing of the quasiparticle matrix element $\langle 0|\sin(\varphi/2)|1\rangle = 0$ (compare with equation (2)). Here the symbol φ represents the dynamical coordinate associated with the operator $\hat{\varphi}$.

effectively providing a fingerprint for the dominant dissipation. To illustrate this idea, in Fig. 1c we represent the expected magnetic flux dependence of the energy relaxation time of the fluxonium qubit, T_1, assuming only quasiparticle dissipation. The T_1 peak around $\Phi_{ext} = 0.5\Phi_0$, where Φ_0 is the flux quantum, is predicted by the coherent cancellation of the dissipative part of equation (1), and is the signature of quasiparticle dissipation in the phase-slip junction[8], so called because its phase difference fluctuates by an amount of the order of 2π.

In Fig. 2a, we present an electron-beam image of an actual device. The fluxonium artificial atom consists of a small Josephson junction providing the nonlinearity, in parallel with the 'superinductor', which shunts charge noise[21]. To suppress charge fluctuations effectively, the superinductor should provide, in the gigahertz domain an impedance larger than the resistance quantum $R_Q = h/(2e)^2 = 6.5\,\mathrm{k\Omega}$, where h is Planck's constant and e is the electron charge. Recent studies[24,25] have demonstrated that superinductors can be reliably implemented using arrays of Josephson junctions. In our design (Fig. 2a), we use an array of 95 junctions, which yields a total inductance of 330 nH, or an equivalent impedance of 20 kΩ at 10 GHz. The relationship between the applied flux and the phase bias is $E_J\sin\varphi + E_L(\varphi - 2\pi\Phi_{ext}/\Phi_0) = 0$, where E_J is the Josephson energy of the phase-slip junction and E_L is the inductive energy of the superinductor.

In general, a superconducting qubit is simultaneously subjected to multiple dissipation channels, including capacitive, inductive, radiative and quasiparticle channels. To observe and identify one particular loss mechanism, in our case quasiparticle loss, we need to increase the characteristic times of all other relaxation channels so as to exceed the quasiparticle-induced T_1. Recent experiments on transmon qubits suggest that this time is expected to be in the millisecond regime[20]. To achieve this goal, we adopted a design strategy similar to the three-dimensional transmon set-up[26]. The sapphire substrate of the fluxonium qubit is clamped in the middle of a copper waveguide cavity (Fig. 2c), which constitutes our sample holder. The coupling to the qubit is mediated by an antenna (Fig. 2b) that couples inductively to the qubit through the shared junctions, and capacitively to the waveguide cavity.

We perform microwave transmission measurements in a circuit quantum electrodynamics set-up[27] (Fig. 2d). The sample holder is thermally anchored to the mixing chamber of a dilution refrigerator with a base temperature of 17 mK. The cavity is housed inside a copper shield coated with infrared-absorbing material, and all microwave lines are filtered above 12 GHz using commercial low-pass filters, circulators and infrared absorbers similar to that in ref. 28. The output line is pre-amplified by a cryogenic high-electron-mobility transistor anchored to the 4 K stage of the cryostat. The copper cavity frequency is $f_c = 8.894$ GHz, with a bandwidth of $\kappa/2\pi = 4$ MHz, and the antenna mode frequency at zero flux is $f_a(0) = 10.6$ GHz. The maximum qubit–cavity dispersive shift is $\chi/2\pi = 0.5$ MHz.

To quantify the junction dissipation, we perform standard time-domain measurements of the energy relaxation time, T_1: after exciting the qubit using a 'saturation' pulse, which divides its population equally between the ground and excited states, we measure the energy decay as a function of time. For any decay channel degree of freedom $\hat{X}$, coupled to the qubit by an operator $\hat{C}$, the relaxation rate can be written, following Fermi's golden rule, as

$$\frac{1}{T_{1X}} = \frac{1}{\hbar^2}\left|\langle 0|\hat{C}|1\rangle\right|^2 S_{II}^X(\omega_{01}) \qquad (2)$$

where $|0\rangle$ and $|1\rangle$ are respectively the ground- and excited-state wavefunctions, $\hbar$ is Planck's constant divided by 2π, S_{II}^X is the current spectral density of noise for the decay channel[29] and ω_{01} is the transition frequency of the qubit:

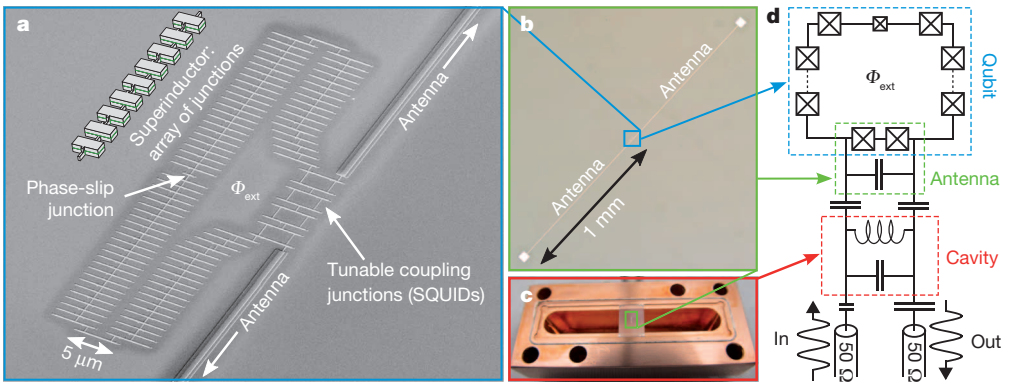


Figure 2 | Experimental set-up. **a**, Electron-beam image of fluxonium sample. The schematic diagram in the upper left corner is a three-dimensional model of the junction array. The Al/AlO_x/Al junctions are made using bridge-free, double-angle evaporation[30]. SQUID, superconducting quantum interference device. **b**, Optical image of the antenna with the fluxonium qubit in the middle. **c**, Photograph of the sapphire chip inside the copper cavity (sample holder). **d**, Electrical scheme for the fluxonium qubit coupling to the input–output microwave lines.

©2014 Macmillan Publishers Limited. All rights reserved

$$S_{II}^{X} = \mathrm{Re}[Y_X(\omega_{01})]\hbar\omega_{01}\left[1+\coth\left(\frac{\hbar\omega_{01}}{2k_B T}\right)\right] \quad (3)$$

Here k_B is Boltzmann's constant.

When the qubit is coupled to dielectric, inductive and radiative loss channels, then $\hat{C}\propto\hat{\varphi}$. Inductive losses in the array would mainly be due to quasiparticles in the array junctions. When the qubit is coupled to quasiparticle dissipation in the phase-slip junction, then $\hat{C}\propto\sin\left(\hat{\varphi}/2\right)$. The dashed lines in Fig. 3a represent numerical evaluations of equation (2) for dielectric loss (green), radiative loss (or 'Purcell' loss; magenta) and quasiparticle loss (red). The continuous black line shows the expected T_1 as a function of applied flux, Φ_{ext}, from the sum of all relaxation rates. Although the qubit is limited by radiative and dielectric loss around $\Phi_{ext}=0$, quasiparticle dissipation becomes dominant as Φ_{ext} increases above $\sim 0.25\Phi_0$. The remarkable, order-of-magnitude, increase in the measured value of T_1 (grey circles) in the vicinity of $\Phi_{ext}/\Phi_0=0.5$ is the unambiguous signature of coherent suppression of quasiparticle dissipation.

The sharpness of the measured T_1 peak at $\Phi_{ext}/\Phi_0=0.5$, which increases by more than an order of magnitude within a flux interval of only $\sim 2\%$ of Φ_0, allows us to place a stringent bound (Methods) on the value of the $\cos\varphi$ prefactor, $\varepsilon>0.99$.

In Fig. 3b, we plot typical energy relaxation curves for different flux biases in the vicinity of the half-flux quantum symmetry point, measured after a saturation pulse. The averaging time for one relaxation curve is 30 min.

A close-up view of the T_1 peak at $\Phi_{ext}/\Phi_0=0.5$ (Fig. 3c) reveals quantitative information about the non-equilibrium quasiparticle density, x_{qp}. The coloured lines correspond to calculations of T_1 for different x_{qp} values. Considering the volume of the superconducting 'islands' on each side of the phase-slip junction ($0.05\,\mu m^3$), we estimate that $x_{qp}=10^{-6}$ corresponds to approximately one quasiparticle per island.

Using the x_{qp} bounds from Fig. 3c, we deduce that the average number of non-equilibrium quasiparticles on the phase-slip junction islands is greater than 0.1 but less than 3. These bounds correspond to measurements taken over a period longer than two months.

As an independent check, we can estimate the average quasiparticle number by analysing individual energy decay curves measured after excitation with a calibrated π-pulse. For a small number of quasiparticles, fluctuations can become important, because the quasiparticle number can change from shot to shot when repeating the measurement. Assuming that the probability, $p_\lambda(n)$, of having n quasiparticles is a Poisson distribution with average λ, $p_\lambda(n)=\lambda^n e^{-\lambda}/n!$, we can then calculate the average time-domain relaxation of the qubit polarization:

$$\langle P(t)\rangle = e^{\lambda\left(\exp\left(-t/\widetilde{T}_{1qp}\right)-1\right)}e^{-t/T_{1R}} \quad (4)$$

Here $\widetilde{T}_{1qp}$ is the relaxation time induced by one quasiparticle, and T_{1R} is the relaxation time associated with the remaining dissipation channel, which in our case is likely to be dielectric loss. We note that, owing to fluctuations in the quasiparticle number, the first factor on the right-hand side of equation (4) is not an exponential decay. Indeed, as we show in Fig. 4a, the fast part of the measured decay is not linear when expressed on a logarithmic scale. The second exponential factor contributes to the linear dependence observed at longer times. Remarkably, the fitted value of the average quasiparticle number, $\lambda=1.1$, falls within the bounds obtained from the above analysis of the data in Fig. 3c.

Because the π-pulse calibration unambiguously provides the signal levels corresponding to the offset and the amplitude of the qubit excitation, plotting the measured data on a logarithmic scale allows us to discriminate between different quasiparticle dynamics. At random moments in time, we measure qubit relaxations compatible with quasi-exponential decays, as shown by the green dashed and magenta solid lines in Fig. 4b. This behaviour can be explained by a reduction, during the measurement

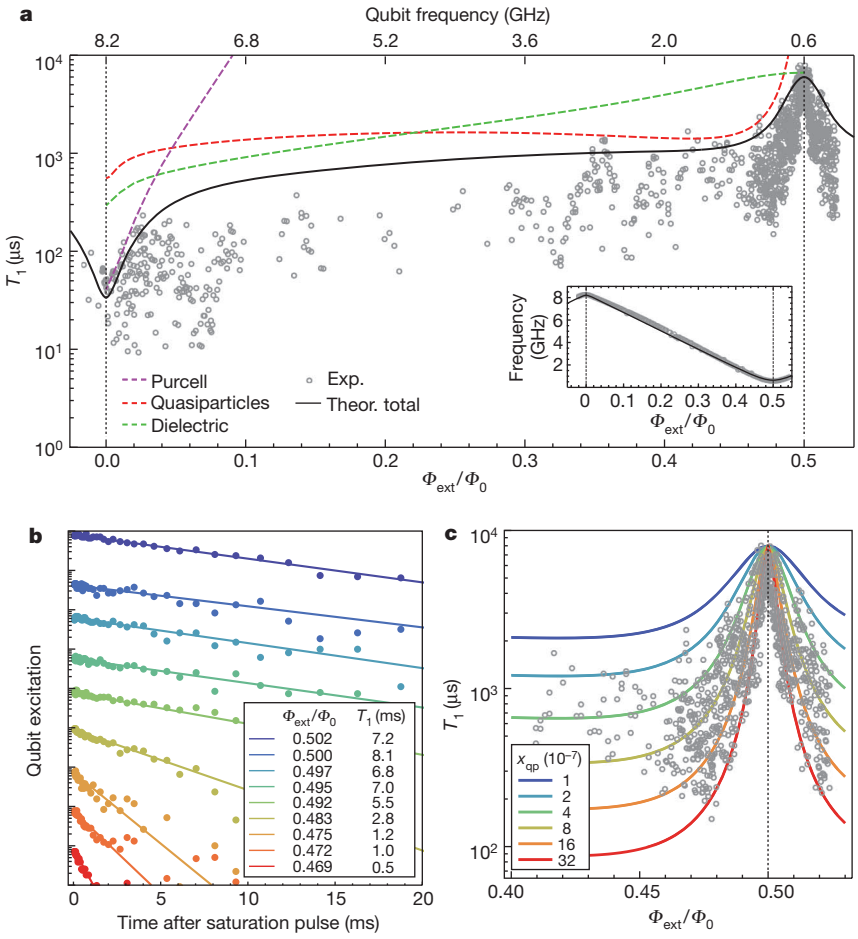


Figure 3 | Flux dependence of the fluxonium qubit T_1 measured by polarization saturation pulses. **a**, The measured values are represented by the grey circles, and the theoretical fit including all sources of dissipation is represented by the black line. The different contributions, T_{1X}, to the total relaxation rate, $1/T_1=\sum 1/T_{1X}$, are plotted as coloured dashed lines. Inductive loss does not significantly contribute to the total decay rate, and the fit lies outside the figure scale (Methods). The inset shows the measured (grey circles) and the theoretical (black line) flux dependence of the $|0\rangle$–$|1\rangle$ qubit transition frequency. **b**, Time-domain measurements of qubit free decay at different flux bias points. The y axis is shown on a logarithmic scale with the traces offset for clarity. The lines represent exponential fits and the inset table lists the fitted T_1 values. **c**, Detailed view in the vicinity of $\Phi_{ext}/\Phi_0=0.5$. The coloured lines show the calculated upper bound on T_1 resulting from the sum of quasiparticle and dielectric dissipation. For this narrow flux interval, we assumed that dielectric loss limits T_1 to a constant value of 8 ms.

©2014 Macmillan Publishers Limited. All rights reserved

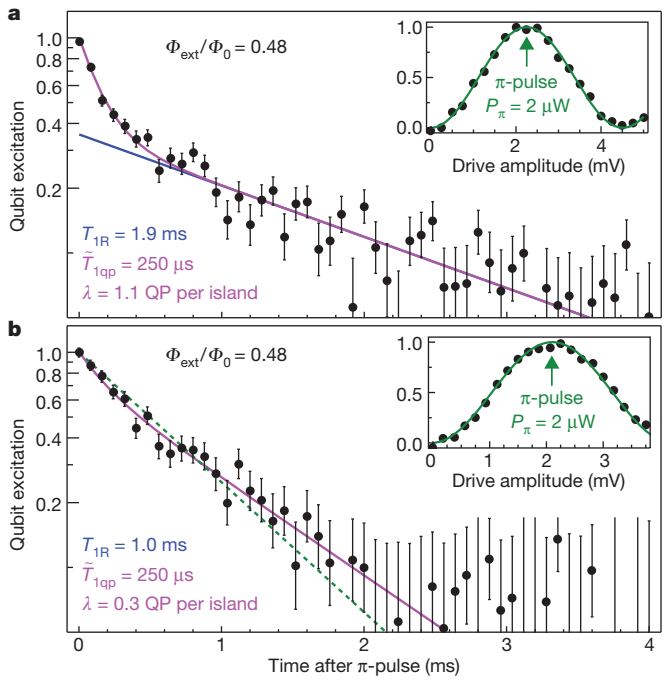


Figure 4 | Time-domain measurements of qubit relaxation after π-pulse excitation. Typical measured traces of non-exponential (**a**) and quasi-exponential (**b**) decays. The magenta curves represent theoretical fits using equation (4). For comparison, the green dashed line in **b** is a fit with a simple exponential decay and time constant of 720 μs. The flux bias point at $\Phi_{ext}/\Phi_0 = 0.48$ is chosen slightly off the maximum of the T_1 peak to allow faster repetition rates for the pulse calibration sequences. The data in **a** and **b** are taken several days apart; the acquisition time for data in each panel is 30 min. We observe non-exponential and quasi-exponential decays with similar occurrence probabilities. The offset and amplitude of the qubit excitation are calibrated using measurements of Rabi oscillations (insets). Error bars, 1 s.d. of the qubit excitation signal.

interval, of the average number of quasiparticles on the central islands ($\lambda = 0.3$), together with a downward fluctuation in T_{1R}. Such fluctuations, illustrated by the scatter of T_1 in Fig. 3a, are evident at any flux.

We have shown that the suppression of quasiparticle dissipation in the vicinity of π-phase bias across the junction confirms Josephson's original prediction[2]. We establish the low-temperature value $\varepsilon = 1$ (expected from Bardeen–Cooper–Schrieffer theory) within less than 1% error for the prefactor in the $1 + \varepsilon\cos\varphi$ interference term. The magnitude of the dissipation indicates the presence of a non-equilibrium population of quasiparticles. Although their origin remains unknown and further characterization of their dynamics goes beyond the scope of this work, the engineering of the qubit susceptibility to quasiparticle loss, as presented here, is a powerful tool for characterizing this ubiquitous dissipation mechanism and ultimately suppressing it altogether. The immediate implications for quantum information processing with superconducting circuits are evident in the achievement of relaxation times well above 1 ms in artificial atoms, an increase by two orders of magnitude.

METHODS SUMMARY

We have measured two nominally identical samples, fabricated in the same lithography cycle. In the main text, we focus on results obtained on sample A, which was measured extensively and exhaustively. In Methods, we show that our results and conclusions were also confirmed by measurements on sample B. Additionally, we describe in detail the microfabrication process, measurement set-up, sample parameters, coherence time measurements and derivations of theoretical expressions used in the main text.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 August 2013; accepted 13 January 2014.

1. Mattis, D. C. & Bardeen, J. Theory of the anomalous skin effect in normal and superconducting metals. *Phys. Rev.* **111,** 412–417 (1958).
2. Josephson, B. D. Possible new effects in superconductive tunnelling. *Phys. Lett.* **1,** 251–253 (1962).
3. Barone, A. & Paternò, G. *Physics and Applications of the Josephson Effect* Ch. 2 (Wiley-VCH, 2005).
4. Clarke, J. & Braginski, A. I. *The SQUID Handbook: Fundamentals and Technology of SQUIDs and SQUID Systems* (Wiley, 2004).
5. Langenberg, D. Physical interpretation of the cos φ term and implications for detectors. *Rev. Phys. Appl.* **9,** 35–40 (1974).
6. Giazotto, F. & Martinez-Perez, M. J. The Josephson heat interferometer. *Nature* **492,** 401–405 (2012).
7. Catelani, G. *et al.* Quasiparticle relaxation of superconducting qubits in the presence of flux. *Phys. Rev. Lett.* **106,** 077002 (2011).
8. Catelani, G., Schoelkopf, R. J., Devoret, M. H. & Glazman, L. I. Relaxation and frequency shifts induced by quasiparticles in superconducting qubits. *Phys. Rev. B* **84,** 064517 (2011).
9. Leppäkangas, J., Marthaler, M. & Schön, G. Phase-dependent quasiparticle tunneling in Josephson junctions: Measuring the cos term with a superconducting charge qubit. *Phys. Rev. B* **84,** 060505 (2011).
10. Beenakker, C. W. J. Universal limit of critical-current fluctuations in mesoscopic Josephson junctions. *Phys. Rev. Lett.* **67,** 3836–3839 (1991).
11. Bretheau, L., Girit, C. O., Pothier, H., Esteve, D. & Urbina, C. Exciting Andreev pairs in a superconducting atomic contact. *Nature* **499,** 312–315 (2013).
12. Harris, R. E. Cosine and other terms in the Josephson tunneling current. *Phys. Rev. B* **10,** 84–94 (1974).
13. Pedersen, N. F., Finnegan, T. F. & Langenberg, D. N. Magnetic field dependence and *Q* of the Josephson plasma resonance. *Phys. Rev. B* **6,** 4151–4159 (1972).
14. Soerensen, O. H., Mygind, J. & Pedersen, N. F. Measured temperature dependence of the cos φ conductance in Josephson tunnel junctions. *Phys. Rev. Lett.* **39,** 1018–1021 (1977).
15. Aumentado, J., Keller, M. W., Martinis, J. M. & Devoret, M. H. Nonequilibrium quasiparticles and 2*e* periodicity in single-Cooper-pair transistors. *Phys. Rev. Lett.* **92,** 066802 (2004).
16. Ferguson, A. J., Court, N. A., Hudson, F. E. & Clark, R. G. Microsecond resolution of quasiparticle tunneling in the single-Cooper-pair transistor. *Phys. Rev. Lett.* **97,** 106603 (2006).
17. Shaw, M. D., Lutchyn, R. M., Delsing, P. & Echternach, P. M. Kinetics of nonequilibrium quasiparticle tunneling in superconducting charge qubits. *Phys. Rev. B* **78,** 024503 (2008).
18. Martinis, J. M., Ansmann, M. & Aumentado, J. Energy decay in superconducting Josephson-junction qubits from nonequilibrium quasiparticle excitations. *Phys. Rev. Lett.* **103,** 097002 (2009).
19. Sun, L. *et al.* Measurements of quasiparticle tunneling dynamics in a band-gap-engineered transmon qubit. *Phys. Rev. Lett.* **108,** 230509 (2012).
20. Ristè, D. *et al.* Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit. *Nature Commun.* **4,** 1913 (2013).
21. Manucharyan, V. E., Koch, J., Glazman, L. I. & Devoret, M. H. Fluxonium: single Cooper-pair circuit free of charge offsets. *Science* **326,** 113–116 (2009).
22. Brooks, P., Kitaev, A. & Preskill, J. Protected gates for superconducting qubits. *Phys. Rev. A* **87,** 052306 (2013).
23. Masluk, N. A. *Reducing the Losses of the Fluxonium Artificial Atom*. PhD thesis, Yale Univ. (2012).
24. Masluk, N. A., Pop, I. M., Kamal, A., Minev, Z. K. & Devoret, M. H. Microwave characterization of Josephson junction arrays: implementing a low loss superinductance. *Phys. Rev. Lett.* **109,** 137002 (2012).
25. Bell, M. T., Sadovskyy, I. A., Ioffe, L. B., Kitaev, A. Y. & Gershenson, M. E. Quantum superinductor with tunable nonlinearity. *Phys. Rev. Lett.* **109,** 137003 (2012).
26. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107,** 240501 (2011).
27. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431,** 162–167 (2004).
28. Santavicca, D. F. & Prober, D. E. Impedance-matched low-pass stripline filters. *Meas. Sci. Technol.* **19,** 087001 (2008).
29. Schoelkopf, R. J., Clerk, A. A., Girvin, S. M., Lehnert, K. W. & Devoret, M. H. in *Quantum Noise in Mesoscopic Physics* (eds Nazarov, Yu. V. & Blanter, Ya. M.) 175–204 (Kluwer Academic, 2002).
30. Lecocq, F. *et al.* Junction fabrication by shadow evaporation without a suspended bridge. *Nanotechnology* **22,** 315302 (2011).

Acknowledgements We acknowledge discussions with L. Frunzio, A. Kamal, N. Masluk and U. Vool. Facilities use was supported by YINQE and NSF MRSEC DMR 1119826. This research was supported by IARPA under grant no. W911NF-09-1-0369, ARO under grant no. W911NF-09-1-0514, the NSF under grants nos DMR-1006060 and DMR-0653377, DOE contract no. DE-FG02-08ER46482 (L.I.G.), and the EU under REA grant agreement CIG-618258 (G.C.).

Author Contributions I.M.P. and K.G. performed the experiment and analysed the data, under the guidance of M.H.D. Theoretical support was provided by G.C. and L.I.G. The experimental design was proposed by I.M.P., K.G., R.J.S. and M.H.D. I.M.P. and M.H.D. led the writing of the manuscript. All authors provided suggestions for the experiment, discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M.P. (ioan.pop@yale.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13184

Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals

Li-Dong Zhao[1], Shih-Han Lo[2], Yongsheng Zhang[2], Hui Sun[3], Gangjian Tan[1], Ctirad Uher[3], C. Wolverton[2], Vinayak P. Dravid[2] & Mercouri G. Kanatzidis[1]

The thermoelectric effect enables direct and reversible conversion between thermal and electrical energy, and provides a viable route for power generation from waste heat. The efficiency of thermoelectric materials is dictated by the dimensionless figure of merit, *ZT* (where *Z* is the figure of merit and *T* is absolute temperature), which governs the Carnot efficiency for heat conversion. Enhancements above the generally high threshold value of 2.5 have important implications for commercial deployment[1,2], especially for compounds free of Pb and Te. Here we report an unprecedented *ZT* of 2.6 ± 0.3 at 923 K, realized in SnSe single crystals measured along the *b* axis of the room-temperature orthorhombic unit cell. This material also shows a high *ZT* of 2.3 ± 0.3 along the *c* axis but a significantly reduced *ZT* of 0.8 ± 0.2 along the *a* axis. We attribute the remarkably high *ZT* along the *b* axis to the intrinsically ultralow lattice thermal conductivity in SnSe. The layered structure of SnSe derives from a distorted rock-salt structure, and features anomalously high Grüneisen parameters, which reflect the anharmonic and anisotropic bonding. We attribute the exceptionally low lattice thermal conductivity (0.23 ± 0.03 W m^{-1} K^{-1} at 973 K) in SnSe to the anharmonicity. These findings highlight alternative strategies to nanostructuring for achieving high thermoelectric performance.

The efficiency of thermoelectric materials and devices is determined by the dimensionless figure of merit (*ZT*), defined as $ZT = (S^2\sigma/\kappa)T$, where S, σ and κ are the Seebeck coefficient, the electrical conductivity and the thermal conductivity, respectively[1,2]. The well-known interdependence of S, σ and κ complicates efforts to develop strategies for improving a material's average *ZT* well above 2.5, especially using less expensive, more Earth-abundant materials[2–4], a feat that could revolutionize the field of thermal energy conversion. Several approaches to enhance *ZT* have emerged in the past decade, including enhancement of Seebeck coefficients (by modifying the band structure[5], heavy valence (conduction) band convergence[6,7], quantum confinement effects[8] and electron energy barrier filtering[9]), and reducing lattice thermal conductivity (by nanostructuring[10] and all-scale hierarchical architecturing[11]) while maintaining hole mobility (band energy alignment between nano-precipitate and matrix[12–14]). Most of these approaches aim to maintain a high power factor (electrical transport properties) and/or reduce the lattice thermal conductivity. Alternatively, high performance could be sought in pristine thermoelectric compounds with intrinsically low thermal conductivity, which may arise from properties such as a large molecular weight[15], a complex crystal structure[16] or charge density wave distortions[17].

We find that SnSe, which is a very stable and simple compound consisting of Earth-abundant elements, exhibits an intrinsically ultralow thermal conductivity. Historically, SnSe was ignored by the thermoelectric community[18,19]; however, its layered and anisotropic crystal structure motivated us to explore its electrical transport properties along all axial directions (unless otherwise noted, all crystallographic directions and planes mentioned in this Letter are defined with respect to its room-temperature form, with *Pnma* space group, #62). Surprisingly, we find that the electrical resistivity is low enough to result in a moderate power factor (along the *b* axis), but, even more surprisingly, we observe that the thermal conductivity of SnSe is intrinsically ultralow (<0.25 W m^{-1} K^{-1} at >800 K), resulting in $ZT = 2.62$ at 923 K along the *b* axis and 2.3 along the *c* axis; these represent the highest *ZT* values reported so far for any thermoelectric system. Along the *a* direction, however, *ZT* is significantly lower, ~0.8. Here, it should be noted that SnSe along the *b* axis shows a room-temperature $ZT = 0.12$, which is comparable to the room-temperature value of 0.15 reported earlier[19]. SnSe, however, reveals high *ZT* values near and above the transition temperature of 750 K at which the structure converts from *Pnma* to *Cmcm*[20–22]. Such ultrahigh *ZT* along two principal directions and the observed crystallographic and *ZT* anisotropy prompted us to investigate the scientific underpinning of these intriguing results.

SnSe adopts a layered orthorhombic crystal structure at room temperature, which can be derived from a three-dimensional distortion of the NaCl structure. The perspective views of the room-temperature SnSe crystal structure along the *a*, *b* and *c* axial directions are shown in Fig. 1a–d. There are two-atom-thick SnSe slabs (along the *b*–*c* plane) with strong Sn–Se bonding within the plane of the slabs, which are then linked with weaker Sn–Se bonding along the *a* direction[20]. The structure contains highly distorted $SnSe_7$ coordination polyhedra, which have three short and four very long Sn–Se bonds, and a lone pair of the Sn^{2+} sterically accommodated between the four long Sn–Se bonds (Fig. 1b). The two-atom-thick SnSe slabs are corrugated, creating a zig-zag accordion-like projection along the *b* axis. The easy cleavage in this system is along the (l00) planes. While cooling from its high-temperature, higher symmetry phase (space group *Cmcm*, #63), SnSe undergoes a displacive (shear) phase transition at ~750–800 K, resulting in a lower symmetry *Pnma* (#62) space group and concomitant matrix transformation of its axes[21,22].

The *ZT* values along the three different crystallographic axes are shown in Fig. 1e. The inset of Fig. 1e shows the SnSe crystals and the cutting directions for typical samples in this study. It should be noted that the *a*, *b* and *c* axial directions were determined through X-ray diffraction (XRD; Extended Data Fig. 1) as well as electron backscatter diffraction (EBSD) analysis (Extended Data Fig. 2).

The electrical conductivities for SnSe crystals along different crystallographic directions show the same temperature-dependent trend (Fig. 2a). We observe three regions: first, metallic transport behaviour from 300 to 525 K; then a change to thermally activated semiconducting behaviour up to ~800 K; and above that, a nearly temperature-independent trend up to 973 K. The first upturn above 525 K is attributed to the thermal excitation of carriers, while the second is related to the phase transition from *Pnma* ($a = 11.49$ Å, $b = 4.44$ Å, $c = 4.135$ Å) to the *Cmcm* ($a = 4.31$ Å, $b = 11.70$ Å, $c = 4.31$ Å) space group[21,22]. It can be readily seen that the electrical conductivities along the *b* and *c* directions are similar, whereas it is lower along the *a* direction. This anisotropy is due to the higher ratio of Hall coefficient to resistivity, R_H/ρ (related to mobility), within the plane of the SnSe slabs than perpendicular to them (that is, along the *a* direction) (Extended Data Fig. 3). The Seebeck

[1]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA. [2]Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA. [3]Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA.

©2014 Macmillan Publishers Limited. All rights reserved

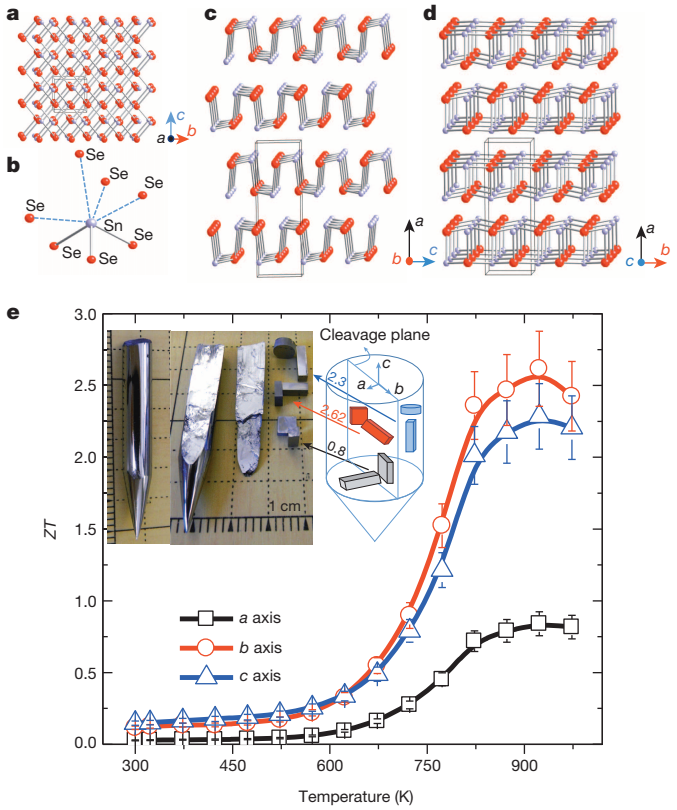


Figure 1 | SnSe crystal structure *Pnma* and *ZT* values. **a**, Crystal structure along the *a* axis: grey, Sn atoms; red, Se atoms. **b**, Highly distorted $SnSe_7$ coordination polyhedron with three short and four long Sn–Se bonds. **c**, Structure along the *b* axis. **d**, Structure along the *c* axis. **e**, Main panel, *ZT* values along different axial directions; the *ZT* measurement uncertainty is about 15% (error bars). Inset images: left, a typical crystal; right, a crystal cleaved along the (l00) plane, and specimens cut along the three axes and corresponding measurement directions. Inset diagram, how crystals were cut for directional measurements; *ZT* values are shown on the blue, red and grey arrows; colours represent specimens oriented in different directions.

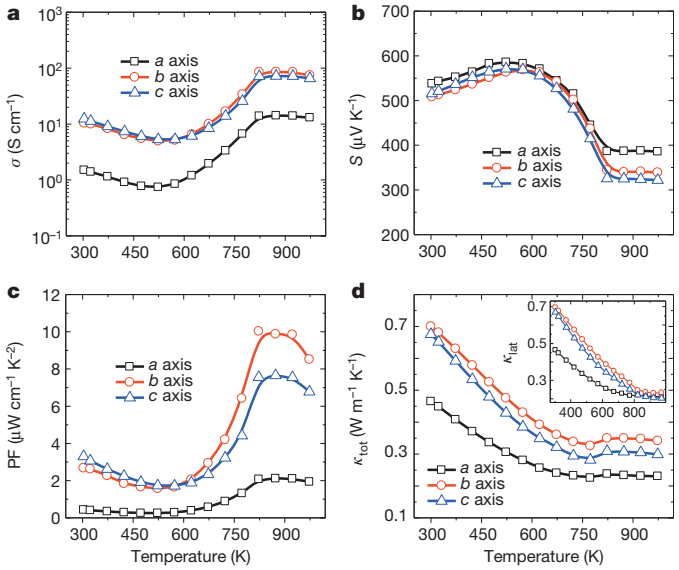


Figure 2 | Thermoelectric properties as a function of temperature for SnSe crystals. **a**, Electrical conductivity. **b**, Seebeck coefficient. **c**, Power factor, PF. **d**, Total thermal conductivity, κ_{tot}. Inset, lattice thermal conductivity, κ_{lat} (same units as κ_{tot}), versus temperature (same units as main panel).

coefficients show almost isotropic behaviour, and are independent of crystallographic directions (Fig. 2b). The gradual decrease of the Seebeck coefficients above 525 K is consistent with the increasing trend in the electrical conductivity, and suggests bipolar conduction and an increasing inverse Hall coefficient, $1/R_H$ (Extended Data Fig. 3a). This behaviour is consistent with our electronic band structure calculations (Extended Data Fig. 4), which show that the bandgap (E_g) decreases considerably from *Pnma* (0.61 eV) to *Cmcm* (0.39 eV), and thus a bipolar conduction process is expected with rising temperature.

The power factor along the *b* axis shows the highest value (10.1 μW $cm^{-1} K^{-2}$) compared to the other two axial directions around 850 K (Fig. 2c); the maximum power factors at 850 K along the *c* and *a* directions are 7.7 μW $cm^{-1} K^{-2}$ and 2.1 μW $cm^{-1} K^{-2}$, respectively. Compared to other state-of-the-art thermoelectrics[2–4], the power factors obtained in SnSe crystals are moderate, but are much higher than those found in other thermoelectrics with intrinsically low overall thermal conductivity (for example, $Yb_{14}MnSb_{11}$, Ag_9TlTe_5, $AgSbTe_2$)[15,16,23]. The highest power factor along *b* is in agreement with the highest R_H/ρ of 250 $cm^2 V^{-1} s^{-1}$ obtained along this axis, which is twofold higher than that along the *c* axis, and tenfold higher than that along the *a* axis (Extended Data Fig. 3b).

The temperature dependence of total thermal conductivity (κ_{tot}) is shown in Fig. 2d. At room temperature (~300 K), the values of κ_{tot} are (in W $m^{-1} K^{-1}$) ~0.46, 0.70 and 0.68 along the *a*, *b* and *c* axis directions, respectively. Compared to state-of-the-art thermoelectrics[2–4], these thermal conductivity values are exceedingly low. Surprisingly, these low values continue to decrease with rising temperature, and at 973 K all fall in the range 0.23–0.34 W $m^{-1} K^{-1}$. The ratio of lattice thermal conductivity (κ_{lat}) to κ_{tot} indicates that κ_{tot} is dominated by phonon transport (Extended Data Fig. 6d). The inset in Fig. 2d indicates that κ_{lat} falls as low as 0.20 W $m^{-1} K^{-1}$ at 973 K along the *a* direction. This is a remarkably low value, which is lower than those obtained even by nanostructuring and all-scale hierarchical architecturing of PbTe-based thermoelectric materials[11].

The dynamic structural behaviour of SnSe above 750 K, involving a reversible phase transition from low-temperature *Pnma* to a high-temperature *Cmcm* space group, helps to preserve the high power factor[22]. This is because the *Cmcm* phase, which is structurally closely related to the *Pnma* phase, exhibits a substantially reduced energy gap and enhanced carrier mobilities while maintaining the ultralow thermal conductivity. We have confirmed this transition using *in situ* sample heating in a transmission electron microscope (TEM). A defect-free lattice image of a SnSe specimen is shown in Fig. 3a, with the inset showing the corresponding selected area diffraction (SAD) pattern taken along the [011] zone axis. It should be noted that kinematically forbidden reflections for the *Pnma* space group, of the type {100}, occur owing to ubiquitous double diffraction in electron diffraction along these directions, and reflect the presence of translational symmetry elements (screw axes/glide planes) in the space group[24]. Other detailed crystal information, including high-resolution TEM images along multiple crystal orientations ([100], [201], [211], [021]) and SAD along [001] and [010], are shown in Extended Data Fig. 7, which confirms the single-crystal SnSe phase at room temperature.

We used the SAD mode (obtained after zone axis alignment with convergent beam electron diffraction) because it is sensitive and can detect subtle changes in crystal symmetry, especially angular rotations. The [211] zone axis at room temperature, which becomes the [121] zone after the transition at about 750–800 K, was chosen to resolve the evolution of the planes (1−1−1) and (0−11). As illustrated by the crystal models in Fig. 3b, viewing along [211] and [121] at room temperature and high temperature, respectively, the plane normal angle between (1−1−1) and (0−11) (marked as blue lines) is expected to change from 86.18° to 89.89° on transition. Figure 3c shows this variation of the angle on the experimentally determined SADs, and we see a good match with the simulated SADs in Extended Data Fig. 8. The diffraction patterns in Fig. 3c were obtained from the same sample area. The sample temperature was

©2014 Macmillan Publishers Limited. All rights reserved

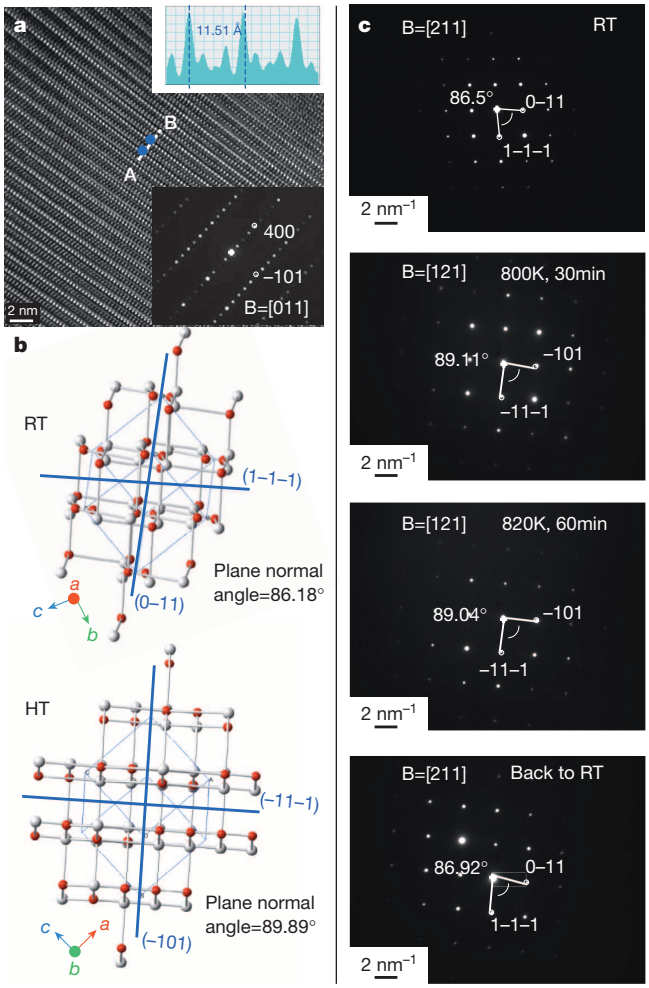


Figure 3 | High-temperature *in situ* TEM observations. **a**, Main panel, high-resolution TEM image of single-crystal SnSe (scale bar, 2 nm). Bottom inset, corresponding diffraction pattern along the [011] zone axis; top inset, the line profile (distance is plotted in Å, *y* axis) along the dotted line AB in the main panel showing the *d* spacing of (100). **b**, Simulated crystal structures of the phase at room temperature (RT; *Pnma*) and at high temperature (HT; *Cmcm*), viewing along the [211] and [121] directions; planes (1−1−1), (−101) and (0−11) are marked by blue lines. **c**, Diffraction patterns obtained at different temperatures. B, zone axis. There is a difference in measured angle between (1−1−1) and (0−11) of about 2.6° between room and elevated temperatures.

increased rapidly from room temperature to 800 K, and then held at this temperature for 30 min after which a diffraction pattern was obtained; the temperature was then increased to 820 K, and the diffraction pattern determined again after holding at this temperature for 60 min; the sample was then brought back to room temperature and the pattern obtained again. The angle between (1−1−1) and (0−11) increased by 2.6° when 800 K was reached. The high-temperature phase is stable up to 820 K for the extended duration used here. There was no noticeable change in terms of *d* spacing and angles between the two SADs at elevated temperatures. In addition, the phase transition was observed to be reversible by returning to room temperature and monitoring the SAD of the *Pnma* phase that reappears. These observations are consistent with a displacive reversible phase transition between *Pnma* and *Cmcm* space groups at about 800 K. Given the subtle angular changes in diffraction observations, it is likely that the phase transition may involve mere shuffling of the SnSe layers, as expected for a displacive second-order phase transition and consistent with the very close relationship of the two structures.

It is known that strong anharmonicity in bonding can give rise to low lattice thermal conductivity in ordered crystal structures[23,25,26]. The strength of the lattice anharmonicity can be estimated from the Grüneisen parameters, which characterize the relationship between phonon frequency and crystal volume change. Therefore, to clarify the origin of the intrinsically low thermal conductivity of SnSe, the phonon and Grüneisen dispersions were calculated using first-principles density functional theory (DFT) phonon calculations within the quasi-harmonic approximation. Figure 4a shows that the acoustic modes along the Γ–X Brillouin zone direction (*a* axis) are significantly softer (with lower Debye temperatures and smaller phonon velocities; see Extended Data Table 1) than those along both the Γ–Y (*b* axis) and the Γ–Z Brillouin zone direction (*c* axis). These soft modes along the *a* axis suggest weak interatomic bonding and possible strong anharmonicity. To quantitatively assess the anharmonicity along the three directions, we plot the dispersion of the Grüneisen parameters of SnSe (Fig. 4b), which shows that the Grüneisen parameters are all very large, with that along the *a* axis being

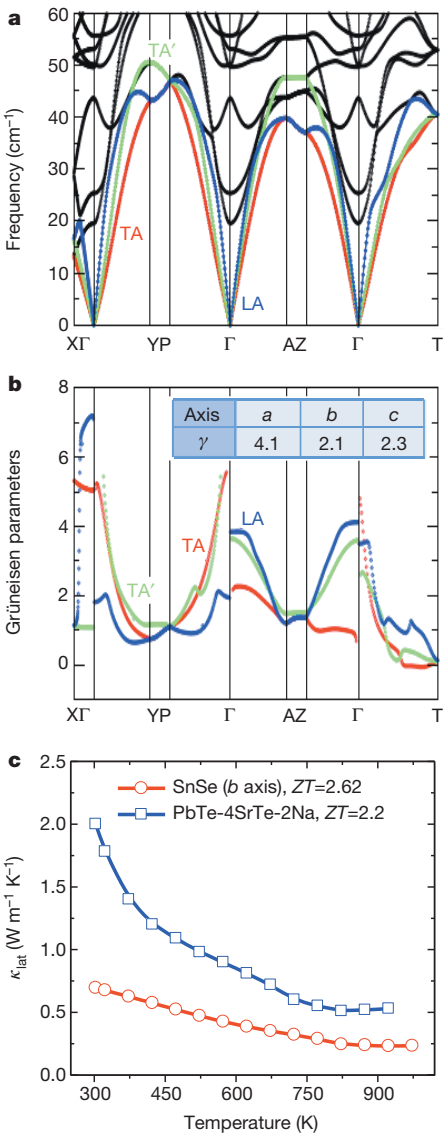


Figure 4 | Theoretically calculated phonon and Grüneisen dispersions, and measured lattice thermal conductivity. **a**, Phonon dispersion. TA, TA′, transverse acoustic phonon scattering branches; LA, longitudinal acoustic phonon scattering branch. **b**, Grüneisen dispersion; inset, the average Grüneisen parameters along *a*, *b* and *c* axes. TA, red colour; TA′, green colour; LA, blue colour. The high symmetry points in the first Brillouin zone can be found in Extended Data Fig. 4. **c**, The lattice thermal conductivity comparison of SnSe along the *b* axis (ZT_{max} = 2.62) and hierarchical architectured PbTe-4SrTe-2Na (ZT_{max} = 2.2)[11].

©2014 Macmillan Publishers Limited. All rights reserved

larger than those along both the *b* and *c* axes. The average Grüneisen parameters along the *a*, *b* and *c* axes are 4.1, 2.1 and 2.3, respectively, as shown in the inset of Fig. 4b. Along the *a* axis, the maximum longitudinal acoustic (LA) Grüneisen parameter around the Γ point is extraordinarily high, ~7.2. In contrast, the Grüneisen parameters are 2.05 for $AgSbTe_2$ (ref. 23), 3.5 for $AgSbSe_2$ (ref. 25) and 1.45 for PbTe (ref. 27), corresponding to measured lattice thermal conductivities at room temperature (in $W\,m^{-1}\,K^{-1}$) of 0.68, 0.48 and 2.4, respectively. The anomalously high Grüneisen parameter of SnSe is a reflection of its crystal structure, which contains very distorted $SnSe_7$ polyhedra (due to the lone pair of Sn^{2+}) and a zig-zag accordion-like geometry of slabs in the *b*–*c* plane. This implies a soft lattice—and if this lattice were mechanically stressed along the *b* and *c* directions, the Sn–Se bond length would not change directly, but instead the zig-zag geometry would be deformed like a retractable spring or an accordion. In addition, along the *a* direction, the weaker bonding between SnSe slabs provides a good stress buffer or 'cushion', thus dissipating phonon transport laterally[28]. The anomalously high Grüneisen parameter is therefore a consequence of the 'soft' bonding in SnSe, which leads to the very low lattice thermal conductivity.

Using our DFT-calculated quantities (Debye temperatures and phonon velocities; see Supplementary Table 1) in the amorphous limit equation[29], we can calculate the minimum lattice thermal conductivity of the three directions (*a*, *b* and *c* axis) at 770 K: $\kappa_{min}{}^{a} = 0.256\ W\,m^{-1}\,K^{-1}$, $\kappa_{min}{}^{b} = 0.360\ W\,m^{-1}\,K^{-1}$ and $\kappa_{min}{}^{c} = 0.326\ W\,m^{-1}\,K^{-1}$. The trend of the theoretically predicted minimum thermal conductivity κ_{min} is in good agreement with the experimental measurements ($\kappa_{lat}{}^{a} < \kappa_{lat}{}^{c} < \kappa_{lat}{}^{b}$). The theoretically calculated minimal thermal conductivities are slightly larger than those from experimental measurements, which could possibly result from (1) the value of the Lorenz number, $1.5 \times 10^{-8}\ V^2\,K^2$, which can vary in the range $(1–2.4) \times 10^{-8}\ V^2\,K^2$; (2) the values of thermal diffusivity, which depend on the details of a fit to time-dependent reflectivity curves (in which instrumental error is about 5%); (3) errors associated with the determined sample thickness, the homogeneity of that thickness, and the determined sample density; and (4) the difficulty of measuring precisely along one crystallographic axis. Indeed, the EBSD analysis shows an ~11° angular deviation of the crystallographic axes from the surface normal (Extended Data Fig. 2).

It is interesting to compare the lattice thermal conductivity of SnSe with that of state-of-the-art thermoelectric systems. The previously reported nanostructured all-scale hierarchical PbTe-4SrTe-2Na (with *ZT* of 2.2) exhibits a lattice thermal conductivity of $0.5\ W\,m^{-1}\,K^{-1}$ (see Fig. 4c)[11]. The unprecedented $ZT \approx 2.62$ of SnSe comes principally from an even lower lattice thermal conductivity of $0.23\ W\,m^{-1}\,K^{-1}$ despite the lack of nanostructuring. We find good experimental repeatability for this high *ZT*, as evidenced by measurements on seven separate crystals prepared independently (Extended Data Fig. 10).

The value of *ZT* (~2.62) at 923 K in SnSe crystals suggests that bulk materials with layered structures, anharmonic bonding and intrinsically ultralow thermal conductivity are promising candidates for providing high thermoelectric performance. It is remarkable that this ultralow thermal conductivity can be realized in a simple compound such as SnSe, as it does not have high molecular weight, a complex crystal structure or a large unit cell. These attributes are generally associated with low thermal conductivity. Compared to other high-performance thermoelectrics, our results for SnSe demonstrate that a high *ZT* also can be realized in simple layered, anisotropic and anharmonic systems, without nanostructuring.

METHODS SUMMARY

The SnSe samples were synthesized first in the form of ingots by heating to 1,223 K over 9.5 h, soaking at that temperature for 6 h, then furnace-cooling to room temperature, followed by vertical Bridgman crystal growth by heating to 1,223 K over 9.5 h, and then cooling from 1,223 K to room temperature with the sample moving at a rate of $2\ mm\,h^{-1}$. The electrical resistivity and Seebeck coefficient were measured simultaneously in a helium atmosphere at 300–973 K using a ULVAC-RIKO ZEM-3 instrument system. We determined carrier concentrations using a custom-made apparatus at 300–823 K. The thermal diffusivity, *D*, was directly measured at 300–973 K by using the laser flash diffusivity method in a commercial Netzsch LFA-457 instrument. The thermal diffusivity was measured along the same direction as the electrical transport. The heat capacity, C_p, was indirectly derived using a representative sample (Pyroceram 9606) in the range 300–973 K. The total thermal conductivity was calculated using the formula $\kappa = DC_p\rho$, where ρ is the sample density, which was determined using the dimensions and mass of the sample and then reconfirmed by measurements using a gas pycnometer (Micromeritics AccuPyc 1340). Transmission electron microscope (TEM) investigation was conducted with a JEOL 2100F at 200 kV. The phonon and Grüneisen dispersions were obtained by using first-principles DFT phonon calculations within the quasi-harmonic approximation.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 November 2013; accepted 24 February 2014.

1. Heremans, J. P., Dresselhaus, M. S., Bell, L. E. & Morelli, D. T. When thermoelectrics reached the nanoscale. *Nature Nanotechnol.* **8,** 471–473 (2013).
2. Zhao, L. D., Dravid, V. P. & Kanatzidis, M. G. The panoscopic approach to high performance thermoelectrics. *Energ. Environ. Sci.* **7,** 251–268 (2014).
3. Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nature Mater.* **7,** 105–114 (2008).
4. Dresselhaus, M. S. *et al.* New directions for low-dimensional thermoelectric materials. *Adv. Mater.* **19,** 1043–1053 (2007).
5. Heremans, J. P. *et al.* Enhancement of thermoelectric efficiency in PbTe by distortion of the electronic density of states. *Science* **321,** 554–557 (2008).
6. Pei, Y. *et al.* Convergence of electronic bands for high performance bulk thermoelectrics. *Nature* **473,** 66–69 (2011).
7. Liu, W. *et al.* Convergence of conduction bands as a means of enhancing thermoelectric performance of *n*-type $Mg_2Si_{1-x}Sn_x$ solid solutions. *Phys. Rev. Lett.* **108,** 166601 (2012).
8. Hicks, L. D. & Dresselhaus, M. S. Effect of quantum-well structures on the thermoelectric figure of merit. *Phys. Rev. B* **47,** 12727–12731 (1993).
9. Heremans, J. P., Thrush, C. M. & Morelli, D. T. Thermopower enhancement in lead telluride nanostructures. *Phys. Rev. B* **70,** 115334 (2004).
10. Hsu, K. F. *et al.* Cubic $AgPb_mSbTe_{2+m}$: bulk thermoelectric materials with high figure of merit. *Science* **303,** 818–821 (2004).
11. Biswas, K. *et al.* High-performance bulk thermoelectrics with all-scale hierarchical architectures. *Nature* **489,** 414–418 (2012).
12. Biswas, K. *et al.* Strained endotaxial nanostructures with high thermoelectric figure of merit. *Nature Chem.* **3,** 160–166 (2011).
13. Zhao, L. D. *et al.* High thermoelectric performance *via* hierarchical compositionally alloyed nanostructures. *J. Am. Chem. Soc.* **135,** 7364–7370 (2013).
14. Zhao, L. D. *et al.* Raising the thermoelectric performance of *p*-type PbS with endotaxial nanostructuring and valence-band offset engineering using CdS and ZnS. *J. Am. Chem. Soc.* **134,** 16327–16336 (2012).
15. Brown, S. R., Kauzlarich, S. M., Gascoin, F. & Snyder, G. J. $Yb_{14}MnSb_{11}$: new high efficiency thermoelectric material for power generation. *Chem. Mater.* **18,** 1873–1877 (2006).
16. Kurosaki, K., Kosuga, A., Muta, H., Uno, M. & Yamanaka, S. Ag_9TlTe_5: a high-performance thermoelectric bulk material with extremely low thermal conductivity. *Appl. Phys. Lett.* **87,** 061919 (2005).
17. Rhyee, J.-S. *et al.* Peierls distortion as a route to high thermoelectric performance in $In_4Se_{3-\delta}$ crystals. *Nature* **459,** 965–968 (2009).
18. Yu, J. G., Yue, A. S. & Stafsudd, O. M. Growth and electronic properties of the SnSe semiconductor. *J. Cryst. Growth* **54,** 248–252 (1981).
19. Wasscher, J. D., Albers, W. & Haas, C. Simple evaluation of the maximum thermoelectric figure of merit, with application to mixed crystals $SnS_{1-x}Se_x$. *Solid-State Electron.* **6,** 261–264 (1963).
20. Peters, M. J. & McNeil, L. E. High-pressure Mossbauer study of SnSe. *Phys. Rev. B* **41,** 5893–5897 (1990).
21. Chattopadhyay, T., Pannetier, J. & Vonschnering, H. G. Neutron-diffraction study of the structural phase-transition in SnS and SnSe. *J. Phys. Chem. Solids* **47,** 879–885 (1986).
22. Baumgardner, W. J., Choi, J. J., Lim, Y.-F. & Hanrath, T. SnSe nanocrystals: synthesis, structure, optical properties, and surface chemistry. *J. Am. Chem. Soc.* **132,** 9519–9521 (2010).
23. Morelli, D. T., Jovovic, V. & Heremans, J. P. Intrinsically minimal thermal conductivity in cubic I-V-VI_2 semiconductors. *Phys. Rev. Lett.* **101,** 035901 (2008).
24. Fultz, B. & Howe, J. M. *Transmission Electron Microscopy and Diffractometry of Materials* (Springer, 2012).
25. Nielsen, M. D., Ozolins, V. & Heremans, J. P. Lone pair electrons minimize lattice thermal conductivity. *Energ. Environ. Sci.* **6,** 570–578 (2013).
26. Zhang, Y. S. *et al.* First-principles description of anomalously low lattice thermal conductivity in thermoelectric Cu-Sb-Se ternary semiconductors. *Phys. Rev. B* **85,** 054306 (2012).
27. Slack, G. A. in *Solid State Physics* (ed. Seitz, F. *et al.*) 1–71 (Academic, 1979).
28. Clarke, D. T. Materials selection guidelines for low thermal conductivity thermal barrier coatings. *Surf. Coat. Technol.* **163–164,** 67–74 (2003).

©2014 Macmillan Publishers Limited. All rights reserved

29. Cahill, D. G., Watson, S. K. & Pohl, R. O. Lower limit to the thermal conductivity of disordered crystals. *Phys. Rev. B* **46,** 6131–6140 (1992).

Acknowledgements This work was supported in part by Revolutionary Materials for Solid State Energy Conversion, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences under award number DE-SC0001054 (L.-D.Z., S.-H.L., Y.Z., H.S., G.T., C.U., C.W., V.P.D. and M.G.K.). TEM work was performed in the (EPIC, NIFTI, Keck-II) facility of the NUANCE Center at Northwestern University. The NUANCE Center is supported by NSF-NSEC, NSF-MRSEC, the Keck Foundation, the State of Illinois and Northwestern University.

Author Contributions L.-D.Z. synthesized the samples, designed and carried out thermoelectric experiments, and wrote the paper. S.-H.L. and V.P.D. performed the TEM characterizations. Y.Z. carried out the calculations. H.S. and C.U. carried out the Hall measurements. G.T. helped with sample synthesis. L.-D.Z., S.-H.L., Y.Z., H.S., G.T., C.U., C.W., V.P.D. and M.G.K. conceived the experiments, analysed the results and co-edited the manuscript. S.-H.L. and Y.Z. contributed equally.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G.K. (m-kanatzidis@northwestern.edu).

©2014 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature13076

Mid-latitude interhemispheric hydrologic seesaw over the past 550,000 years

Kyoung-nam Jo[1], Kyung Sik Woo[2], Sangheon Yi[1], Dong Yoon Yang[1], Hyoun Soo Lim[3], Yongjin Wang[4], Hai Cheng[5,6] & R. Lawrence Edwards[6]

An interhemispheric hydrologic seesaw—in which latitudinal migrations of the Intertropical Convergence Zone (ITCZ) produce simultaneous wetting (increased precipitation) in one hemisphere and drying in the other—has been discovered in some tropical and subtropical regions[1–3]. For instance, Chinese and Brazilian subtropical speleothem (cave formations such as stalactites and stalagmites) records show opposite trends in time series of oxygen isotopes (a proxy for precipitation variability) at millennial to orbital timescales[2,3], suggesting that hydrologic cycles were antiphased in the northerly versus southerly subtropics. This tropical to subtropical hydrologic phenomenon is likely to be an initial and important climatic response to orbital forcing[3]. The impacts of such an interhemispheric hydrologic seesaw on higher-latitude regions and the global climate system, however, are unknown. Here we show that the antiphasing seen in the tropical records is also present in both hemispheres of the mid-latitude western Pacific Ocean. Our results are based on a new 550,000-year record of the growth frequency of speleothems from the Korean peninsula, which we compare to Southern Hemisphere equivalents[4]. The Korean data are discontinuous and derived from 24 separate speleothems, but still allow the identification of periods of peak speleothem growth and, thus, precipitation. The clear hemispheric antiphasing indicates that the sphere of influence of the interhemispheric hydrologic seesaw over the past 550,000 years extended at least to the mid-latitudes, such as northeast Asia, and that orbital-timescale ITCZ shifts can have serious effects on temperate climate systems. Furthermore, our result implies that insolation-driven ITCZ dynamics may provoke water vapour and vegetation feedbacks in northern mid-latitude regions and could have regulated global climate conditions throughout the late Quaternary ice age cycles.

The Korean peninsula in northeast Asia is situated on the boundary between the world's largest continent and largest ocean[5] and has a typical mid-latitude climate that is strongly affected by the East Asian monsoon (EAM) system (Fig. 1). The peninsula has experienced dramatic palaeogeographic changes and eustatic sea-level changes during the late Quaternary glacial–interglacial cycles. For example, during the last glacial maximum (LGM), vast areas of the Yellow Sea (about 500,000 km^2) to the west of the peninsula were completely exposed subaerially, preventing

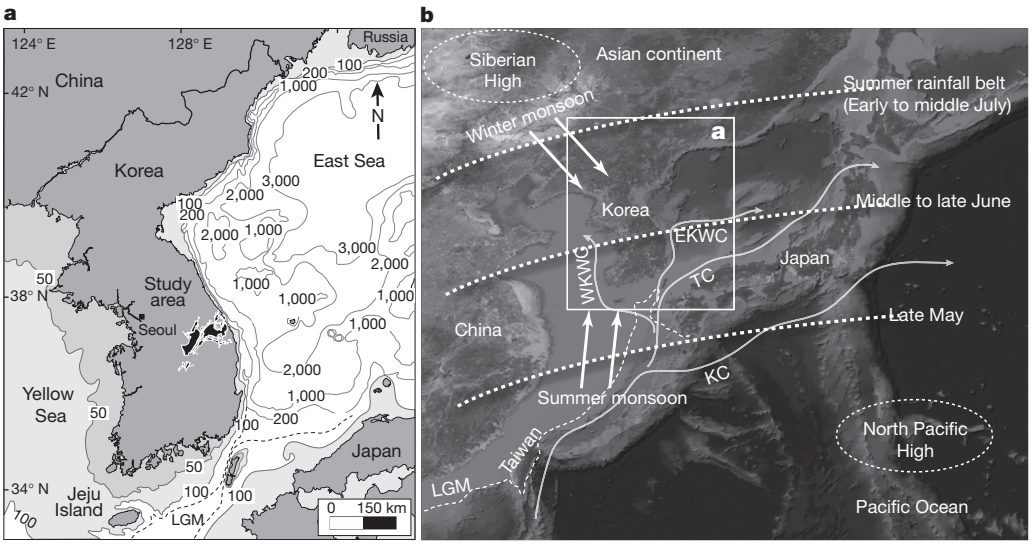


Figure 1 | Location of the study area and current atmospheric and oceanographic conditions. Also shown are coastlines during the last glacial maximum (LGM; dashed lines). **a**, Black-shaded areas represent the location of the Cambro–Ordovician Joseon Supergroup, where we collected all the speleothem samples. The black lines indicate the current coastlines, and grey lines with numbers (depths in metres) show current bathymetric features. A vast area of the Yellow Sea was subaerially exposed during the LGM[6] (shaded areas and dashed lines). **b**, The arrows describe atmospheric and oceanographic conditions. KC, Kuroshio current; TC, Tsushima current; EKWC, East Korean warm current; WKWC, West Korean warm current. Thick dashed lines indicate the current summer rainfall belt, called Jangma in Korea. The background image is from NASA's Visible Earth Web Site (http://visibleearth.nasa.gov/).

[1]Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, South Korea. [2]Department of Geology, Kangwon National University, Gangwondo 200-701, South Korea. [3]Department of Geological Sciences, Pusan National University, Busan 609-735, South Korea. [4]College of Geography Science, Nanjing Normal University, Nanjing 210097, China. [5]Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an 710049, China. [6]Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA.

©2014 Macmillan Publishers Limited. All rights reserved

warm current inflow[6] (Fig. 1). These palaeoenvironmental settings produced major changes in terrestrial ecosystems during the late Quaternary glacial–interglacial cycles[7,8], and the changes on the Korean peninsula can be considered to reflect global changes in vegetation composition and density. Proxy records from the Korean peninsula may provide invaluable information on continent–ocean climate linkages and the impacts of global changes on ice sheets and sea level.

However, only a few orbital-scale proxy records representing more than one glacial–interglacial cycle have been reported, possibly owing to the lack of directly datable materials. In this study, we have applied the ^{238}U–^{234}U–^{230}Th dating technique to describe the first record of speleothem growth frequency from the northeast Asian sector over the past 550,000 years (550 kyr).

The modern annual mean temperature (AMT) of the study area is approximately 10–13 °C, with strong seasonal variations of >30 °C between January and August[9]. The winter temperature is generally below freezing, with monthly mean temperatures of −6 °C to −7 °C in January. Annual mean precipitation (AMP) is approximately 1,100–1,400 mm, with high seasonality. Three quarters of the total annual precipitation falls during the humid summer season from June to September. All the caves in this study were formed in Cambro–Ordovician carbonate rocks of the Joseon Supergroup, which is composed mainly of limestone, with a minor contribution of dolomite and argillaceous limestone (Extended Data Table 1). The caves we studied were mostly formed along numerous joint and fault planes created by Jurassic tectonic movements[10] (Fig. 1). The atmospheric environment in Korean limestone caves is characterized by stable temperature and humidity (see Methods for the detailed discussion) (Extended Data Fig. 1). Typical limestone caves in the Korean peninsula contain large numbers of actively growing speleothems[11], indicating that current climatic and environmental conditions are suitable for the formation of various types of speleothem.

We obtained 148 ^{230}Th ages from 24 speleothems (22 stalagmites and two flowstones from 15 typical limestone caves) using multi-collector inductively coupled plasma mass spectroscopic techniques (Extended Data Fig. 2 and Supplementary Table 1). Filtering to avoid Holocene bias removed 22 dates (Extended Data Fig. 3). We supplemented the plot of the number of speleothem depositions with textural examination results and confirmed hiatuses (Extended Data Figs 4 and 5). Our speleothem growth frequency data were augmented by 640 new and previous stable isotope data from three stalagmites (ED1, GE1 and DY1 from Eden cave, Gwaneum cave and Daeya cave, respectively).

The growth frequency of Korean speleothems has varied widely over time (Fig. 2 and Extended Data Fig. 6), indicating the past fluctuations of speleothem growth in the limestone caves of the Korean peninsula. Although various high-resolution speleothem $\delta^{18}O$ records from tropical to subtropical China have been reported, this record is the first to show speleothem growth frequency from the northeast Asian temperate region. Each high peak provides the precise intervals for the active growth periods of the speleothems, and these data can be directly correlated with orbital-scale changes in global palaeoclimate records closely related to advances and retreats of massive continental ice sheets. High- and low-growth-frequency phases of Korean speleothems are well matched to interglacial/interstadial and glacial/stadial periods, respectively, based on marine isotope stages (MIS)[12] (Fig. 2).

Although the older part of the record was more or less unclear compared with the penultimate glacial period, it is notable that Korean speleothems grew markedly in interglacial/interstadial periods (MIS 1, MIS 3, MIS 5a, MIS 5c, MIS 5e, MIS 7a, MIS 7c, MIS 7e, MIS 9, MIS 11 and MIS 13), whereas their growth significantly declined during glacial/stadial conditions (MIS 2, MIS 4, MIS 6, MIS 7d, MIS 8 and MIS 12). Moreover, almost all the peaks in our record are possibly linked to the second half of each peak of Northern Hemisphere summer insolation[13] and the EAM intensity record[14,15], as well as to interglacial/interstadial periods represented by isotopic records from an Antarctic ice core[16] and Devils Hole[17] (19 of 27 peaks in Northern Hemisphere summer insolation, 13 of 16 peaks in marine and terrestrial records), with a few exceptions mainly before 400 kyr ago (Fig. 2). These agreements between the growth frequency record of the Korean speleothems and the major palaeoclimatic sequences strongly suggest that Korean speleothem growth was primarily controlled by climatic and environmental changes. This

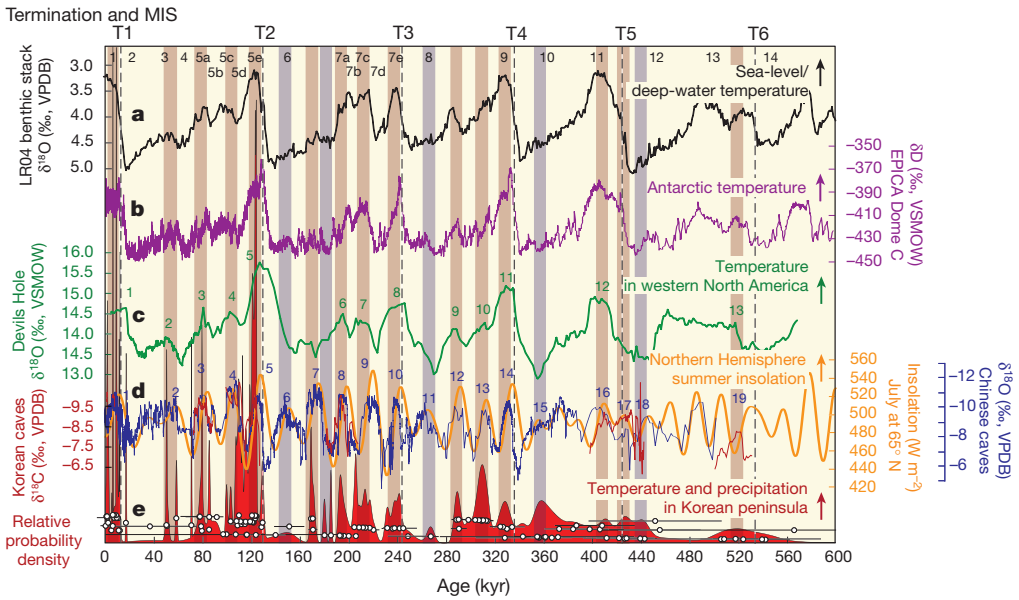


Figure 2 | Comparison of the growth frequency record from Korean speleothems with major palaeoclimatic records over the last 600 kyr. **a**, $\delta^{18}O$ record of LR04 benthic stack[12]. The marine isotope stage (MIS) is indicated above the curve. Vertical dashed lines show the timing of glacial terminations. **b**, δD record of EPICA Dome C, Antarctica[16]. **c**, $\delta^{18}O$ record of vein calcite from Devils Hole, western USA[17]. The numbers above the curve represent high peak intervals and are matched to the Korean speleothem record. **d**, Summer insolation at 65° N (ref. 13) superimposed on the composite $\delta^{18}O$ record from the Hulu, Sanbao and Linzhu caves in China[15]. The numbers above the curve represent high peaks matched to the Korean speleothem record. **e**, Growth frequency record of Korean speleothems with the simple plot of ^{230}Th dates (open circles with horizontal lines). Error bars of each date depict 2σ error. See the Methods for a detailed description of the sample collection. Periods of high growth frequency are shown in brown (interglacial/interstadial) and purple (glacial) vertical bars. The $\delta^{13}C$ records (red curves) closely overlap with Chinese $\delta^{18}O$ records and are from three stalagmites: GE-1, DY-1 and ED-1. The scale of the $\delta^{13}C$ values is that of ED-1. The GE-1 and DY-1 records were tuned to the Chinese $\delta^{18}O$ records.

©2014 Macmillan Publishers Limited. All rights reserved

result is in accordance with the growth of a speleothem in northeastern China (Shihua cave, 39° 47′ N, 115° 56′ E) which shows that much faster growth occurred during warmer periods[18].

Indeed, the most basic and simplest palaeoclimatic clue provided by speleothems is whether the speleothems grew under the prevailing climatic conditions existing outside the caves[19]. In theory, speleothem growth requires both a sufficient supply of cave water and suitable temperature, that is, at least positive effective precipitation[2,4] and ice- and permafrost-free conditions[20,21]. Additionally, for speleothem growth in the temperate climatic regime, productive vegetation and soil depths thick enough to produce and store soil CO_2 and moisture should be present[4,19]. Thus, we attributed the high speleothem growth frequency during the interglacial (and/or pluvial) periods to reinforcement of surficial vegetation productivity and soil depth under clearly favourable climatic conditions. This interpretation is confirmed by noticeably lower $\delta^{13}C$ values (about −10.7‰ to −9.0‰ in the Eden cave scale) from three stalagmites during interglacial and interstadial periods than during glacial conditions (about −8.5‰ to −6.5‰) (Fig. 2), suggesting that terrestrial productivity in the study area was highly intensified during interstadial and interglacial periods. For the case of glacial periods with low speleothem growth, our interpretation is also supported by pollen and vertebrate evidence[7,8] from the Korean peninsula as well as numerical model simulations. Considering both the significantly reduced growth frequency of the Korean speleothems and the estimated amount of precipitation during rainy seasons, the peninsula probably experienced a further southward shift or even failure of the Jangma rain belt (also called the Meiyu in Japan and Bai-u in China), which is the strong frontal system that develops each year in early summer between the maritime Okhotsk and North Pacific air masses (see Methods for a detailed discussion).

The timing of palaeoclimate changes in our record is correlated with boreal monsoonal changes (Fig. 2). Movement of the ITCZ—which is the main controlling factor of interhemispheric monsoon intensity—has climatic impacts in middle and higher latitudes, as successive displacements in zonal flow and related air masses compensate for the ITCZ motion, as in modern seasonal movements[22,23]. The Korean peninsula, as a mid-latitude temperate region and one of the northernmost parts of the EAM sector, should be highly sensitive to even small shifts in the mean position and vitality of the ITCZ (Fig. 1). Impacts of ITCZ migration on regions at latitudes higher than the tropics and subtropics can be ascertained by the linkage between the Korean speleothem record and Mediterranean sapropel events[24] (Fig. 3). Our record shows excellent matches with the sapropel events, indicating that both regions have been affected by synchronous movement of the ITCZ, which resulted in atmospheric circulation changes in both regions. This observation in the mid-latitudes is also shown in speleothem evidence of millennial-scale fluctuations from the southwestern United States[22] and from southwestern New Zealand[23]. Thus, our data support the idea that ITCZ shift causes latitudinal displacement of all climatic zones.

To delineate the mid-latitudinal interhemispheric impacts of ITCZ migration, we combined the growth frequency record of Korean speleothems with the Southern Hemisphere geographical equivalent from Naracoorte, southeastern Australia[4] (Fig. 3). The Korean speleothem record shows a distinct inverse correlation with the Southern Hemisphere mid-latitude speleothem record. For example, the Holocene and last interglacial periods are indicated as remarkable growth periods of Korean speleothems but as periods of little growth of Australian speleothems. Similarly, the southeastern Australian speleothems actively developed during cool interstadials to stadials, whereas Korean speleothems showed only moderate growth near interstadial peaks. Notably, no or very low growth frequency was shown in both hemispheres during the massive glacial periods including MIS 4 and MIS 7d (Fig. 3).

Once again, this interhemispheric anti-correlation supports the theory that variations in the mean position of the ITCZ, directly affected by solar insolation, redistributed the global configuration of latitudinal climatic zones and air masses. During peak interglacial periods, the Korean peninsula had been influenced by a relatively warmer and wetter climate than during interstadials, owing to the enhanced effect of the North Pacific high-pressure system along with the northern position of the ITCZ (Figs 1 and 3)[5]. In contrast, during cool interstadials to stadials, southeastern Australia has experienced a wetter and cooler climate by southward displacement of the ITCZ and/or the strong effect of Walker circulation[4]. Alternatively, a steeper temperature gradient between the southern polar front and ITCZ could have induced stronger Southern Hemisphere westerly winds and prevailing southern frontal systems along with early Equatorward advance of the polar front and

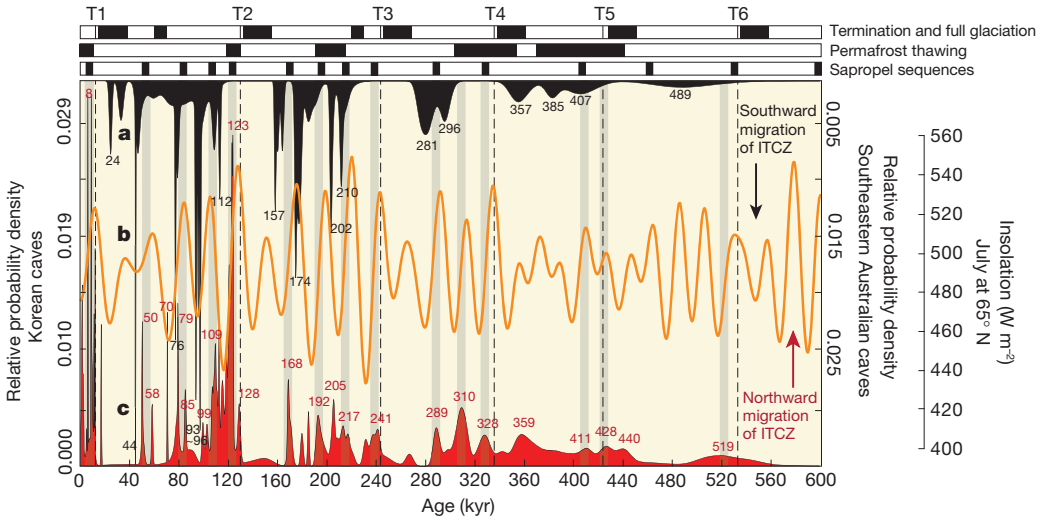


Figure 3 | Comparison of the Korean speleothem record with the Southern Hemisphere counterpart record from southeastern Australia. **a**, Growth frequency record of southeastern Australian speleothems[4]. The black numbers around each peak show the age. **b**, Summer insolation at 65° N (ref. 13). **c**, Growth frequency record of Korean speleothems. The red numbers around each peak show the age of growth periods. The light grey vertical bars represent the periods of high growth frequency indicated in the Korean speleothem record during interglacial/interstadial periods. We note the distinctive anti-correlations between the northern and southern mid-latitude speleothem records. The vertical dashed lines show the timings of glacial terminations. The bars in the uppermost part of the figure indicate periods of massive glaciation, including MIS 4 and MIS 7d. The middle bar shows periods of permafrost thawing from the eastern Siberian speleothem record[21]. Mediterranean sapropel events reported by ref. 24 and other literature are shown in the lower bar.

©2014 Macmillan Publishers Limited. All rights reserved

expansion of the sea ice extent[23]. This inference describes the mid-latitude interhemispheric hydrologic seesaw from peak interglacial to mid-sized glacial periods. Because the southeastern Australian record reflects effective precipitation (higher precipitation with cooler temperature in Naracoorte)[4], this is also likely to indicate synchronous ice-driven forcing in both the northern and southern mid-latitudes[25] in the western Pacific sector. During full glacial conditions, very limited growth frequencies were found for speleothems in both hemispheres (Fig. 3). This suggests that both regions experienced maximal cold and dry conditions as the northern continental ice sheet and southern sea ice reached their maximum extents. In other words, this indicates the equatorward push of the ITCZ by strong ice-driven forcing and weakening of the global rainfall system in both hemispheres[26] (Extended Data Fig. 7).

The interconnecting atmospheric processes described above are likely to be concurrent with the climatic feedbacks in the northern mid-latitude regions (Extended Data Fig. 7). Insolation-driven northward displacement of the ITCZ can bring about sensible heating and positive feedback by the potent effect of water vapour as a major greenhouse gas[27] in the northern mid- to high-latitude regions. It is possible that this type of positive feedback was combined with the precipitation–vegetation–albedo feedback[28,29], leading to acceleration in the increase in surface air temperature and rapid ice melting in the Northern Hemisphere, especially during glacial terminations (see Methods for the detailed discussion). Our record strongly supports this explanation because the higher speleothem growth frequency with lower $\delta^{13}C$ values in the Korean peninsula should have been accompanied by a significant increase in vegetation density and/or soil depth during the terminations (Fig. 2). According to pollen records of the Korean peninsula[8], the enhanced monsoon precipitation and increasing temperature led to shifts in vegetation types from glacial grassland to interglacial forest. This explanation for the vegetation-related feedback mechanism in the northern mid-latitude regions is also in line with a recent numerical model simulation[30].

Given that the northernmost region of the Korean peninsula is thought to have been one of the southern limits of discontinuous permafrost in the Asian continent during glacial maxima, our hypothesis could be closely related to thawing of the Northern Hemispheric permafrost region, a massive reservoir of greenhouse gases. Recently, the speleothem data from Eastern Siberia reveals that permafrost thawing is highly sensitive to slight increases in global temperature and can be potentially coupled with substantial release of greenhouse gases trapped in it into the atmosphere[21]. But this greenhouse effect caused by large-scale permafrost thawing is likely to be most effective during glacial terminations because the magnitude of the thawing during other periods, such as the ends of stadials, must be significantly reduced owing to the inadequate extension of permafrost areas at those time periods.

METHODS SUMMARY

We obtained 148 available uranium-series dates using a multi-collector inductively coupled plasma mass spectrometer (MC-ICP-MS; Thermo-Finnigan Neptune) at the Minnesota Isotope Laboratory, University of Minnesota, USA. All age error bars in this paper depict 2σ errors. Seventeen of the dates were previously published[31,32]. The relative probability density curve for the total dates of the Korean speleothems was plotted using Isoplot/Ex 3.0 software (http://bgc.org/isoplot_etc/isoplot.html). This plot has an inherent defect that may be artificially caused by the sampling strategy (closer sub-sampling intervals in a particular part). To solve this problem, we supplemented the plot of the number of speleothem depositions with textural examination results and confirmed hiatuses. Another weakness of the plot was that Holocene speleothem samples were more common than older samples. To address this, we primarily focused on the timings of peaks in the probability density curve, which were regarded as the periods of high-frequency speleothem growth in the Korean peninsula.

Oxygen and carbon isotope ratios were analysed at the Isotope Laboratory at Nanjing Normal University, China. In total, 640 measurements of $\delta^{18}O$ and $\delta^{13}C$ were made on the three speleothems using a Finnigan MAT-253 ratio mass spectrometer with an on-line, automated, carbonate preparation system (Kiel III). Isotopic data of 346 measurements from GE1 and DY1 were published. Measured C- and O-isotope ratios in per mil (‰) notation are reported relative to the belemnite from the Pee Dee Belemnite formation standard (PDB). Analytical errors for $\delta^{18}O$ and $\delta^{13}C$ are 0.15‰ and 0.2‰, respectively. Detailed age models for the time-series isotopic records of the stalagmites were derived by linear interpolation between all dates. The final age models for the three stalagmites were tuned to fit Chinese cave $\delta^{18}O$ records within age errors.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 June 2013; accepted 22 January 2014.
Published online 30 March 2014.

1. Severinghaus, J. P. Monsoons and meltdowns. *Science* **326,** 240–241 (2009).
2. Wang, X. *et al.* Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature* **432,** 740–743 (2004).
3. Cruz, F. W. Jr *et al.* Insolation-driven changes in atmospheric circulation over the past 116,000 years in subtropical Brazil. *Nature* **434,** 63–66 (2005).
4. Ayliffe, L. K. *et al.* 500 ka precipitation record from southeastern Australia: evidence for interglacial relative aridity. *Geology* **26,** 147–150 (1998).
5. Nakagawa, T. *et al.* Regulation of the monsoon climate by two different orbital rhythms and forcing mechanisms. *Geology* **36,** 491–494 (2008).
6. Kim, J.-M. & Kucera, M. Benthic foraminifer record of environmental changes in the Yellow Sea (Hwanghae) during the last 15,000 years. *Quat. Sci. Rev.* **19,** 1067–1085 (2000).
7. Lim, H. S. *et al.* Vertebrate burrows in late Pleistocene paleosols at Korean Palaeolithic sites and their significance as a stratigraphic marker. *Quat. Res.* **68,** 213–219 (2007).
8. Yi, S. & Kim, S.-J. Vegetation changes in western central region of Korean Peninsula during the last glacial (ca. 21.1–26.1 cal kyr BP). *Geosci. J.* **14,** 1–10 (2010).
9. Korea Meteorological Administration. http://web.kma.go.kr/eng/biz/climate_01.jsp (2011).
10. Chough, S. K. *et al.* Tectonic and sedimentary evolution of the Korean peninsula: a review and new view. *Earth Sci. Rev.* **52,** 175–235 (2000).
11. Jo, K., Woo, K. S., Hong, G. H., Kim, S. H. & Suk, B. C. Rainfall and hydrological controls on speleothem geochemistry during climatic events (droughts and typhoons): an example from Seodong Cave, Republic of Korea. *Earth Planet. Sci. Lett.* **295,** 441–450 (2010).
12. Lisiecki, L. E. & Raymo, M. E. A. Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20,** PA1033 (2005).
13. Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10,** 297–317 (1991).
14. Wang, Y. *et al.* Millennial- and orbital-scale changes in the East Asian monsoon over the past 224,000 years. *Nature* **451,** 1090–1093 (2008).
15. Cheng, H. *et al.* The climatic cyclicity in semiarid-arid central Asia over the past 500,000 years. *Geophys. Res. Lett.* **39,** L01705 (2012).
16. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317,** 793–796 (2007).
17. Winograd, I. J. *et al.* Continuous 500,000-year climate record from vein calcite in Devils Hole, Nevada. *Science* **258,** 255–260 (1992).
18. Tan, M. *et al.* Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophys. Res. Lett.* **30,** 1617 (2003).
19. Baker, A., Smart, P. L. & Ford, D. C. Northwest European palaeoclimate as indicated by growth frequency variations of secondary calcite deposits. *Palaeogeogr., Palaeoecol. Palaeoclimatol.* **100,** 291–301 (1993).
20. Spötl, C., Mangini, A., Frank, N., Eichstädter, R. & Burns, S. J. Start of the last interglacial period at 135 ka: evidence from a high Alpine speleothem. *Geology* **30,** 815–818 (2002).
21. Vaks, A. *et al.* Speleothems reveal 500,000-year history of Siberian permafrost. *Science* **340,** 183–186 (2013).
22. Asmerom, Y., Polyak, V. J. & Burns, S. J. Variable winter moisture in the southwestern United States linked to rapid glacial climate shifts. *Nature Geosci.* **3,** 114–117 (2010).
23. Whittaker, T. E., Hendy, C. H. & Hellstrom, J. C. Abrupt millennial-scale changes in intensity of Southern Hemisphere westerly winds during marine isotope stages 2–4. *Geology* **39,** 455–458 (2011).
24. Ziegler, M., Tuenter, E. & Lourens, L. J. The precession phase of the boreal summer monsoon as viewed from the eastern Mediterranean (ODP Site 968). *Quat. Sci. Rev.* **29,** 1481–1490 (2010).
25. Weber, M. E. *et al.* Interhemispheric ice-sheet synchronicity during the Last Glacial Maximum. *Science* **334,** 1265–1269 (2011).
26. Stager, J. C., Ryves, D. B., Chase, B. M. & Pausata, F. S. R. Catastrophic drought in the Afro-Asian monsoon region during Heinrich Event 1. *Science* **331,** 1299–1302 (2011).
27. Soden, B. J., Wetherald, R. T., Stenchikov, G. L. & Robock, A. Global cooling after the eruption of Mount Pinatubo: a test of climate feedback by water vapor. *Science* **296,** 727–730 (2002).
28. Kutzbach, J., Bonan, G., Foley, J. & Harrison, S. P. Vegetation and soil feedbacks on the response of the African monsoon to orbital forcing in the early to middle Holocene. *Nature* **384,** 623–626 (1996).
29. Cheddadi, R. & Bar-Hen, A. Spatial gradient of temperature and potential vegetation feedback across Europe during the late Quaternary. *Clim. Dyn.* **32,** 371–379 (2009).
30. McCarthy, M. P. *et al.* The influence of vegetation on the ITCZ and South Asian monsoon in HadCM3. *Earth Syst. Dyn.* **3,** 87–96 (2012).

©2014 Macmillan Publishers Limited. All rights reserved

31. Jo, K. *et al.* Textural and carbon isotopic evidence of monsoonal changes recorded in a composite-type speleothem from Korea since MIS 5a. *Quat. Res.* **74,** 100–112 (2010).
32. Jo, K. *et al.* Holocene and Eemian climatic optima in the Korean Peninsula based on textural and carbon isotopic records from the stalagmite of the Daeya Cave, South Korea. *Quat. Sci. Rev.* **30,** 1218–1231 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Kangwon National University Cave Investigation Club (KNUCIC) for collecting some of the speleothem samples, S. S. Lee for the supplementary statistical test and K. R. Ludwig for providing the analytical software package. This research was a part of the project titled K-IODP (KIGAM; Korea Institute of Geoscience and Mineral Resources) funded by the Ministry of Oceans and Fisheries, Korea. This project was also partially supported by Basic Research Project (GP2009-005) of KIGAM, grants NSFC 41230524 and NBRP 2013CB955902 (to H.C.) and US NSF grants 1103403 and 1337693 (to R.L.E. and H.C.).

Author Contributions K.-n.J. collected most of the information and wrote the first draft of the manuscript. K.S.W. discussed the palaeoclimatic implications of the data and improved the manuscript. S.Y. commented on the palaeoclimatic interpretations. D.Y.Y. and H.S.L. provided research grants to K.S.W. and K.-n.J. for obtaining data on caves and speleothems. Y.W. provided the stable isotope data, and H.C. and R.L.E. supplied some of the $^{230}Th/^{234}U$ dating results. All authors approved the submitted form of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.S.W. (wooks@kangwon.ac.kr).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13086

A new fossil species supports an early origin for toothed whale echolocation

Jonathan H. Geisler[1], Matthew W. Colbert[2] & James L. Carew[3]

Odontocetes (toothed whales, dolphins and porpoises) hunt and navigate through dark and turbid aquatic environments using echolocation; a key adaptation that relies on the same principles as sonar[1]. Among echolocating vertebrates, odontocetes are unique in producing high-frequency vocalizations at the phonic lips, a constriction in the nasal passages just beneath the blowhole, and then using air sinuses and the melon to modulate their transmission[2,3]. All extant odontocetes seem to echolocate[2,4]; however, exactly when and how this complex behaviour—and its underlying anatomy—evolved is largely unknown. Here we report an odontocete fossil, Oligocene in age (approximately 28 Myr ago), from South Carolina (*Cotylocara macei*, gen. et sp. nov.) that has several features suggestive of echolocation: a dense, thick and downturned rostrum; air sac fossae; cranial asymmetry; and exceptionally broad maxillae. Our phylogenetic analysis places *Cotylocara* in a basal clade of odontocetes, leading us to infer that a rudimentary form of echolocation evolved in the early Oligocene, shortly after odontocetes diverged from the ancestors of filter-feeding whales (mysticetes). This was followed by enlargement of the facial muscles that modulate echolocation calls, which in turn led to marked, convergent changes in skull shape in the ancestors of *Cotylocara*, and in the lineage leading to extant odontocetes.

Cetacea
Odontoceti
Xenorophidae
Cotylocara macei gen. et sp. nov.

Etymology. *Cotylo*, for cavity, and *cara*, for head, both Greek, in reference to bilateral pits on the frontals. Species name recognizes Mace Brown for his contributions to the founding of a Natural History Museum at the College of Charleston, South Carolina, USA.

Holotype. CCNHM-101. Nearly complete skull, partial dentaries, three cervical vertebrae (C2, C3 or C4, C6), and portions of at least seven ribs found in close association and presumed to represent a single individual (Figs 1 and 2 and Extended Data Figs 1, 2 and 3).

Locality and age. Drainage ditch in the College Park Subdivision, Berkeley County, South Carolina. Approximate coordinates: 33° 1′ 50′′ N,

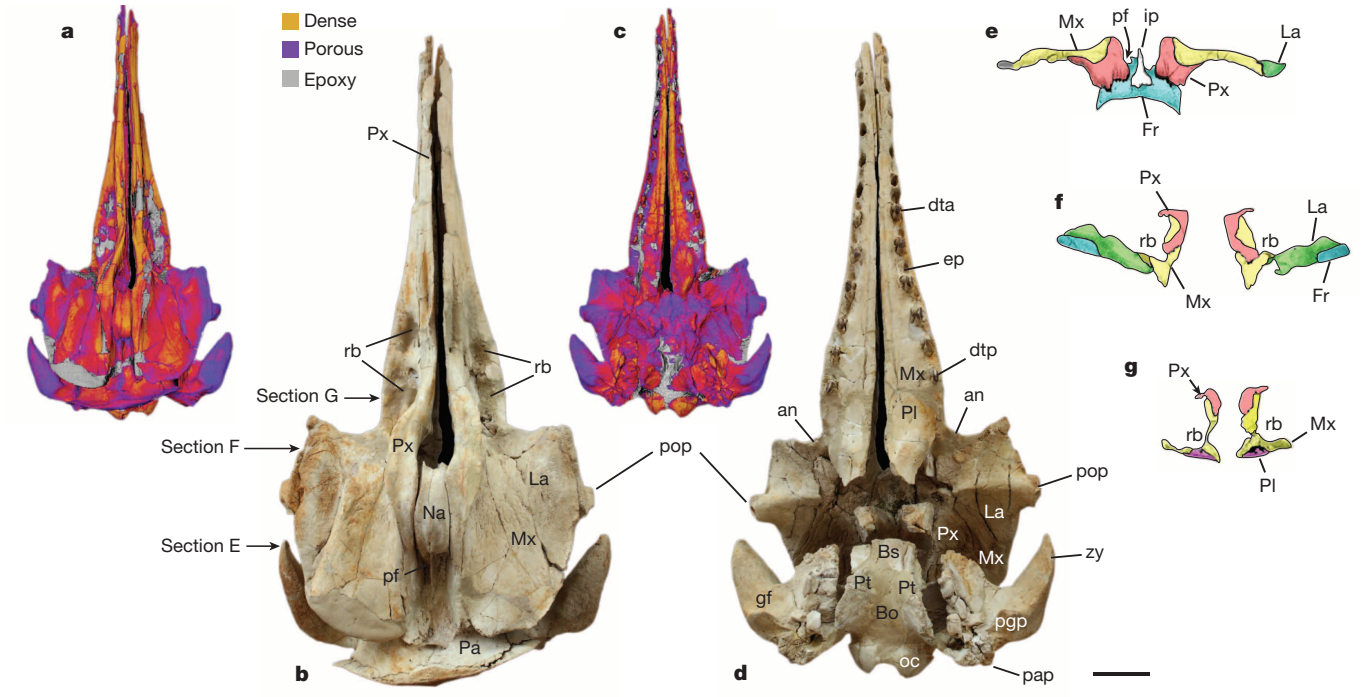


Figure 1 | Holotype skull of *Cotylocara macei* (CCNHM-101), including cross-sections and bone density. **a**, **b**, CT model and photograph of skull in dorsal view. **c**, **d**, Same as **a**, **b**, but in ventral view. **e–g**, CT cross-sections of skull; locations indicated in **b**. Colours in **a** and **c** reflect the relative density of bone, whereas in **e–g**, each bone is a separate colour as follows: blue, frontal; green, lacrimal; purple, palatine; red, premaxilla; white, interparietal; yellow, maxilla. an, antorbital notch; Bo, basioccipital; Bs, basisphenoid; dta, anteriormost double-rooted tooth; dtp, posteriormost double-rooted tooth; ep, embrasure pit; Fr, frontal; gf, glenoid fossa; ip, interparietal; La, lacrimal; Mx, maxilla; Na, nasal; oc, occipital; Pa, parietal; pap, paroccipital process; pf, postnarial fossa; pgp, postglenoid process; Pl, palatine; pop, postorbital process; Pt, pterygoid; Px, premaxilla; rb, rostral basin; zy, zygomatic process. Scale bar is 5 cm.

[1]Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Northern Boulevard, Old Westbury, New York 11568, USA. [2]Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas 78712, USA. [3]Department of Geology and Environmental Geosciences, Natural History Museum, College of Charleston, Charleston, South Carolina 29424, USA.

©2014 Macmillan Publishers Limited. All rights reserved

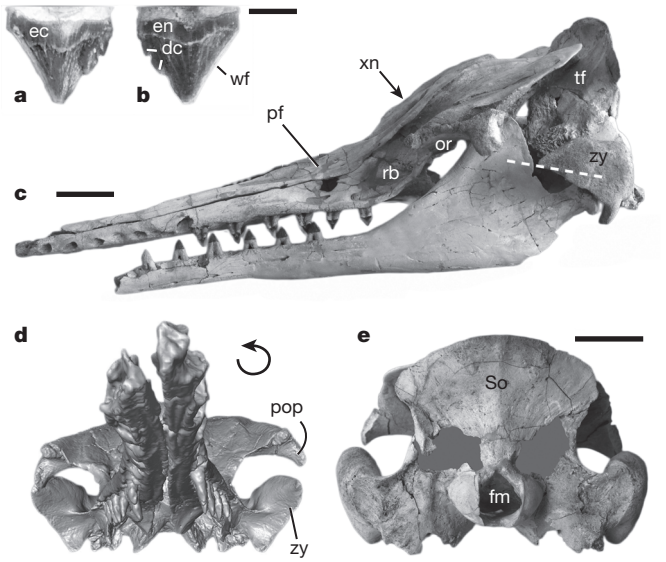


Figure 2 | Holotype skull and dentition of *Cotylocara macei* (CCNHM-101). **a**, **b**, Last maxillary tooth from the left side of the skull in labial and lingual views. **c**, Skull in lateral view, downturned to indicate angle between rostrum and basicranial stem (marked by dashed line). **d**, CT-generated model of skull in anterior view with perspective showing counterclockwise torsion between the face and braincase. **e**, Posterior view of skull. dc, denticles; ec, ectocingulum; en, entocingulum; fm, foramen magnum; or, orbit; pf, premaxillary fossa; pop, postorbital process; rb, rostral basin; So, supraoccipital; tf, temporal fossa; wf, wear facet; xn, external nares; zy, zygomatic process. Scale bar for **a** and **b** is 5 mm, scale bars for **c** and **e** are 5 cm.

80° 5′ 51.4′′ W. Bed 2 of the Chandler Bridge Formation, late Oligocene in age[5].

Diagnosis. The species differs from all known cetaceans (whales, dolphins, porpoises) in that it has deep, postnarial fossae on frontals that are separated by a median septum of the interparietal (Fig. 1b, e) and a vertical supraoccipital whose anterior face is overlapped by parietals (Figs 1b and 2). Like the xenorophids *Albertocetus meffordorum* and *Xenorophus sloani*, the lateral portion of the frontal is covered by the ascending process of the lacrimal, the frontal has a window on its ventral side that exposes strips of lacrimal and maxilla (Extended Data Fig. 4), the premaxilla underlies the ascending process of the maxilla (Fig. 1e), bilateral rostral basins are present, and the petrosal (Extended Data Fig. 2) has an elongate lateral tuberosity that articulates with the squamosal[6] (not preserved in *Xenorophus*). However, *Cotylocara* differs from these taxa, as well as *Archaeodelphis patrius*, in having premaxillae that overhang maxillae (Fig. 1f); maxillae overhanging the squamosal fossae; thick asymmetrical nasals with transversely compressed crests at their anterolateral corners; and anterolaterally projecting, dorsoventrally deep zygomatic processes of the squamosals[6] (the last two characters are not preserved in *Xenorophus*).

In addition to the phonic lips, odontocete echolocation is associated with a suite of soft-tissue facial structures, including a fat-filled body called the melon, which is thought to focus vocalizations[7]; a series of pneumatic diverticulae that allow for near continuous vocalizations while diving and may act as acoustic reflectors (for example, premaxillary sinus, inferior vestibule)[8]; and a complex of muscles that change the size and/or shape of the melon, the diverticulae, and nasal passages[9]. Although fossils of these soft tissues have not been found, we contend that some aspects of these tissues can be inferred from fossil skulls. For example, *C. macei* has a pair of large, rostral basins immediately ventrolateral to the premaxillary sac fossae that breach the lateral walls of the infraorbital canals (Extended Data Fig. 4), similar to exposure of the mandibular division of the trigeminal nerve by the pterygoid sinus in extant cetaceans. We suggest that expansion of the premaxillary air sinus excavated the rostral basin (Extended Data Fig. 5). Although shallow or flat premaxillary sac 'fossae' have been identified in other Oligocene odontocetes[10,11], *C. macei* represents the first strong osteological evidence for air sinuses anterior to the bony nares in basal odontocetes. Similarly, we interpret the deep postnarial fossae as being excavated by air sinuses, a hypothesis supported by the thin median septum that resembles the bony lamina in the pterygoid sinuses of many odontocete skulls[12]. We tentatively suggest that these fossae accommodated a diverticulum of the inferior vestibule, possibly a homologue to the caudal sac of phocoenids[13]. In addition, the rostral basins and the postnarial fossae are formed by low-density bone, as in the premaxillary sac fossa of *Tursiops* (Fig. 1 and Extended Data Fig. 6).

Cotylocara resembles extant odontocetes in having an expanded maxilla that covers most of the supraorbital process of the frontal, and this expansion may be related to the fact that it serves as the proximal attachment for the maxillonasolabialis muscle and homologues[10,14]. In odontocetes the maxillonasolabialis is hypertrophied, split into several different muscles, and inserts onto the melon, nasal passages, and nasal diverticula[2,9,13,15]. If the site of attachment for the maxillonasolabialis is conserved, then a larger muscle would be coupled with a larger maxilla. It is generally assumed that some of the homologues of the maxillonasolabialis in odontocetes play key roles in the production of sound at the phonic lips[13], particularly because the nasal muscles in mysticetes, which do not echolocate, are considerably smaller and less complex[16].

Most extant odontocetes have asymmetric heads where parts of the 'median' plane are shifted to the left side and some structures on the right side are much larger than those on the left[9]. Some researchers have suggested that this asymmetry is involved in the production of high-frequency vocalizations[9,15]; for example, one nasal passage might be primarily for vocalizations whereas the other only for respiration[17]. However, some studies suggest that both the right and left phonic lips generate sounds[3], and others suggest that cranial asymmetry might allow for large prey to be swallowed whole[18]. *Cotylocara* represents one of the earliest and most definitive examples of odontocete cranial asymmetry. The left rostral basin is 140% wider and longer than its counterpart on the right (Fig. 1b); the premaxillae are asymmetric where they border the rostral basins and external bony nares; and portions of the internasal and interparietal sutures are to the left of the median plane. A recent study[19] identified directional asymmetry in the skulls of archaeocetes, the paraphyletic stem group of crown cetaceans, and speculated that this asymmetry was an adaptation for using high-frequency sound to locate prey, as has been documented in owls[20]. In archaeocetes the internasal suture is shifted towards the right side, and in anterior view, the rostrum is twisted counterclockwise[19], although clockwise rotation occurs in a skull of *Cynthiacetus*[21]. Interestingly, in *Cotylocara* the rostrum is also rotated counterclockwise (Fig. 2b), but some sutures are shifted to the left, not the right. The rostral torsion is probably not the result of post-mortem deformation because the Chandler Bridge Formation is unconsolidated and flat-lying[5]. Whether the unique pattern of asymmetry in *Cotylocara* is indicative of high-frequency sound production, sound reception, or the size of its prey can only be determined once the basis for asymmetry in extant odontocetes is better understood.

Computed tomography (CT) scans reveal that the premaxillae of *C. macei* are dense, including where they encircle the external bony nares and overhang the maxillae (Fig. 1a). A similarly dense and overhanging premaxilla occurs in Cuvier's beaked whale (*Ziphius cavirostris*), and finite element modelling supports the hypothesis that the dense premaxilla acts as an acoustic reflector in that taxon[22]. Likewise, the density contrast in *C. macei* could reflect much of the sound forward, particularly because the rostrum is deflected 20° downwards relative to the basicranial stem (Fig. 2 and Extended Data Fig. 1). The nasals and dorsal portions of the maxillae and lacrimals are also osteosclerotic, raising the possibility that the large interorbital shield could be an acoustic reflector.

Taken together, the cranial features described above make a compelling case that *Cotylocara* could echolocate, and that has important implications for the origin of this complex behaviour. Our phylogenetic analysis

©2014 Macmillan Publishers Limited. All rights reserved

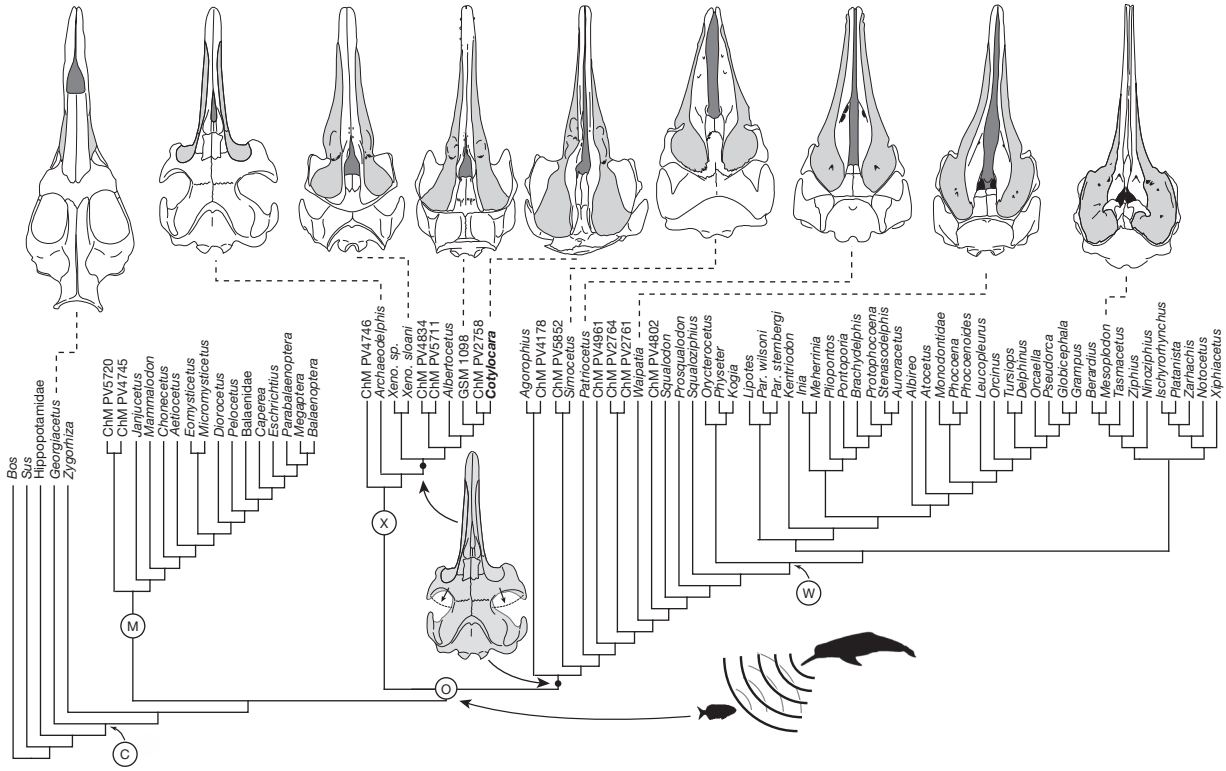


Figure 3 | Phylogeny supported by the present study, with evolution of echolocation and skull shape. Skulls are not to scale; maxilla is in grey. Solid circles indicate convergent evolution of the roof to the temporal fossa (see Supplementary Information). Aspects of the skull of *Archaeodelphis* were reconstructed, braincase of *Xenorophus sloani* based on CCNHM-168, and three skulls redrawn from published figures[10,29,30], adapted with permission from ref. 10 (Smithsonian Institution Scholarly Press) and ref. 29 (San Diego Society of Natural History). C, Cetacea; CCNHM, College of Charleston Natural History Museum, South Carolina; ChM PV, The Charleston Museum, Vertebrate Paleontology Collection, South Carolina; GSM, Georgia Southern Museum; M, Mysticeti; O, Odontoceti; *Par.*, *Parapontoporia*; X, Xenorophidae; *Xeno.*, *Xenorophus*; W, crown Odontoceti.

of a modified supermatrix of morphological and molecular data[23–25] resulted in one most parsimonious tree (Fig. 3). *Cotylocara* is nested in Xenorophidae, the first clade to branch from the odontocete stem, and is most closely related to *Albertocetus meffordorum* among the described taxa. If *Cotylocara* could echolocate, it is most parsimonious to infer that the ability to produce sounds at the phonic lips evolved in the common ancestor of all known odontocetes, shortly after they diverged from mysticetes (that is, 34 to 30 Myr ago[26]). Our results are consistent with earlier hypotheses, based on less evidence, that the early Oligocene odontocetes *Simocetus rayi*[10] and two unnamed species[11] could echolocate.

Like other studies[6,27], we found *Archaeodelphis patrius* to be a basal xenorophid, which is significant because the frontals and maxillae in that taxon do not roof over the temporal fossa. When xenorophids are arranged according to our phylogenetic tree, there is a trend towards posterior migration of the frontals, maxillae and premaxillae. This trend occurred in parallel along the stem of crown Odontoceti (Fig. 3, Extended Data Fig. 7 and Supplementary Information). As a result, *Cotylocara* looks superficially more like some crown odontocetes (for example, *Mesoplodon*) than it resembles basal xenorophids. Similarly, cranial asymmetry is likely to have evolved twice because basal xenorophids, like *Archaeodelphis* and *Xenorophus*, lack it. An extreme interpretation of these patterns would be that echolocation evolved separately in xenorophids, but we prefer the simpler explanation that a rudimentary form of echolocation was present in the common ancestor of xenorophids and other odontocetes, and this shared behaviour facilitated the convergent cranial evolution in both clades. In support of this view, we corroborated that the lateral expansion of the maxillae over the orbits is a synapomorphy of Odontoceti[14,15,26,27], suggesting that hypertrophy of the maxillonasolabialis and echolocation evolved before cranial convergence. Our hypothesis predicts that the inner ear of basal odontocetes should be specialized to hear the echoes of high-frequency sounds produced at the phonic lips. Although the inner ear labyrinth of *Cotylocara* is not well preserved, anatomical studies on the inner ear of other xenorophids should provide a robust test for an early evolution of odontocete echolocation.

METHODS SUMMARY

Our phylogeny (Fig. 3) is the single most parsimonious tree (41155.88 steps in length) for a supermatrix[25] (311 morphological and 60,851 molecular characters) that was modified by adding *Cotylocara*, *Albertocetus meffordorum*, an undescribed xenorophid (GSM 1098), and a single morphological character (see Supplementary Information). The tree was obtained using the computer application TNT[28], with default search parameters under a 'New Technology Search'. As in previous studies[24,25], between-character scaling was achieved by down-weighting ordered, multistate, morphological characters so that they have the same minimum length as a binary character.

The density model of the skull was based on CT data from the Medical University of South Carolina (cubic voxel size is 0.584 mm per side) and was generated in Avizo v.7.1, with colours assigned according to grayscale values. Reconstructed portions were digitally isolated before applying the density-coded colour palette. Current density should be a reasonable proxy for original density because exposed breaks show no permineralization and the Chandler Bridge Formation is unlithified.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 October 2013; accepted 29 January 2014.
Published online 12 March 2014.

1. Au, W. W. L. *The Sonar of Dolphins* (Springer-Verlag, 1993).
2. Cranford, T. W., Amundin, M. & Norris, K. S. Functional morphology and homology in the odontocete nasal complex: implications for sound generation. *J. Morphol.* **228,** 223–285 (1996).
3. Cranford, T. W. *et al.* Observation and analysis of sonar signal generation in the bottlenose dolphin (*Tursiops truncatus*): evidence for two sonar sources. *J. Exp. Mar. Biol. Ecol.* **407,** 81–96 (2011).

©2014 Macmillan Publishers Limited. All rights reserved

4. Ketten, D. R. in *Marine Mammal Sensory Systems* (eds Thomas, J.A., Kastelein, R.A. & Supin, A.Y.) 53–75 (Plenum Press, 1992).
5. Sanders, A. E., Weems, R. E., Lemon, E. M. & Jr.. Chandler Bridge Formation—a new Oligocene stratigraphic unit in the lower coastal plain of South Carolina. *US Geol. Surv. Bull.* **1529,** 105–124 (1982).
6. Uhen, M. D. A new *Xenorophus*-like odontocete cetacean from the Oligocene of North Carolina and a discussion of the basal odontocete radiation. *J. Syst. Palaeontology* **6,** 433–452 (2008).
7. McKenna, M. F., Cranford, T. W., Berta, A. & Pyenson, N. Morphology of the odontocete melon and its implications for acoustic function. *Mar. Mamm. Sci.* **28,** 690–713 (2012).
8. Reidenberg, J. S. & Laitman, J. T. Sisters of the sinuses: cetacean air sacs. *Anat. Rec (Hoboken)* **291,** 1389–1396 (2008).
9. Mead, J. G. Anatomy of the external nasal passages and facial complex in the Delphinidae (Mammalia: Cetacea). *Smithson. Contrib. Zool.* **207,** 1–35 (1975).
10. Fordyce, R. E. in *Cenozoic Mammals of Land and Sea: Tributes to the Career of Clayton E. Ray* (ed. Emry, R. J.) 185–222 (Smithsonian Institution Scholarly Press, 2002).
11. Barnes, L. G., Goedert, J. L. & Furusawa, H. in *Western Association of Vertebrate Paleontologists with Mesa Southwest Museum and Southwest Paleontological Society: First Meeting of the New Millenium* Vol. 8 (eds McCord, R. D. & Boaz, D.) 91–100 (Southwest Paleontological Society, 2001).
12. Fraser, F. C. & Purves, P. E. Hearing in cetaceans, evolution of the accessory air sacs and the structure and function of the outer and middle ear in recent cetaceans. *Bull. Brit. Mus. Nat. Hist. Zool.* **7,** 1–140 (1960).
13. Huggenberger, S., Rauschmann, M. A., Vogl, T. J. & Oelschläger, H. H. Functional morphology of the nasal complex in the harbor porpoise (*Phocoena phocoena* L.). *Anat. Rec. (Hoboken)* **292,** 902–920 (2009).
14. Fordyce, E. & de Muizon, C. in *Secondary Adaptation of Tetrapods to Life in Water* (eds Mazin, J.-M. & de Buffrénil, V.) 169–234 (Verlag Dr. Friedrich Pfeil, 2001).
15. Heyning, J. E. Comparative facial anatomy of beaked whales (Ziphiidae) and a systematic revision among the families of extant Odontoceti. *Nat. Hist. Mus. Los Angeles Cty. Contrib. Sci.* **405,** 1–64 (1989).
16. Carte, A. & Macalister, A. On the anatomy of *Balaenoptera rostrata. Phil. Trans. R. Soc. Lond.* **158,** 201–261 (1868).
17. Madsen, P. T., Wisniewska, D. & Beedholm, K. Single source sound production and dynamic beam formation in echolocating harbor porpoises (*Phocoena phocoena*). *J. Exp. Biol.* **213,** 3105–3110 (2010).
18. Macleod, C. D. *et al.* Breaking symmetry: the marine environment, prey size, and the evolution of asymmetry in cetacean skulls. *Anat. Rec. (Hoboken)* **290,** 539–545 (2007).
19. Fahlke, J. M., Gingerich, P. D., Welsh, R. C. & Wood, A. R. Cranial asymmetry in Eocene archaeocete whales and the evolution of directional hearing in water. *Proc. Natl Acad. Sci. USA* **108,** 14545–14548 (2011).
20. Norberg, R. A. in *Ecology and Conservation of Owls* (eds Newton, I., Kavanagh, R., Olsen, J. & Taylor, I.) 329–342 (Csiro Publishing, 2002).
21. Martínez-Cáceres, M. & de Muizon, C. A new basilosaurid (Cetacea, Pelagiceti) from the late Eocene to early Oligocene Otuma Formation of Peru. *C. R. Palevol* **10,** 517–526 (2011).
22. Cranford, T. W., Krysl, P. & Hildebrand, J. A. Acoustic pathways revealed: simulated sound transmission and reception in Cuvier's beaked whale (*Ziphius cavirostris*). *Bioinspir. Biomim.* **3,** 016001 (2008).
23. McGowen, M. R., Spaulding, M. & Gatesy, J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol. Phylogenet. Evol.* **53,** 891–906 (2009).
24. Geisler, J. H., McGowen, M. R., Yang, G. & Gatesy, J. A supermatrix of genomic, morphological, and paleontological data from crown Cetacea. *BMC Evol. Biol.* **11,** 112 (2011).
25. Geisler, J. H., Godfrey, S. J. & Lambert, O. A new genus and species of late Miocene inioid (Cetacea, Odontoceti) from the Meherrin River, North Carolina, U.S.A. *J. Vertebr. Paleontol.* **32,** 198–211 (2012).
26. Gatesy, J. *et al.* A phylogenetic blueprint for a modern whale. *Mol. Phylogenet. Evol.* **66,** 479–506 (2013).
27. Geisler, J. H. & Sanders, A. E. Morphological evidence for the phylogeny of Cetacea. *J. Mamm. Evol.* **10,** 23–129 (2003).
28. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24,** 774–786 (2008).
29. Fordyce, R. E. in *Contributions in Marine Mammal Paleontology Honoring Frank Whitmore Jr* (eds Berta, A. & Deméré, T. A.) 147–176 (San Diego Society of Natural History, 1994).
30. Dubrovo, I. A. & Sanders, A. E. A new species of *Patriocetus* (Mammalia, Cetacea) from the late Oligocene of Kazakhstan. *J. Vertebr. Paleontol.* **20,** 577–590 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by funds from the New York Institute of Technology, College of Charleston, and M. Brown. During this project we benefitted from discussions with B. Beatty, P. Gingerich, W. Hillenius, M. Mihlbachler, A. Sanders and N. Solounias. We thank A. Sanders and The Charleston Museum for access to specimens under their care. B. Dulguun drew the skulls in Fig. 3. We acknowledge the Deptartment of Radiology, Medical University of South Carolina, and T. Holden in particular, for conducting CT scans of *Cotylocara*. We thank T. Rowe for access to the *Tursiops* scan data, which were scanned using his NSF digital libraries initiative grant (IIS-987-4781).

Author Contributions J.H.G. designed the research plan and conducted the phylogenetic analysis. M.W.C. generated three-dimensional models from CT data and conducted related analyses on this data. J.H.G. wrote the paper with discussion from J.L.C and M.W.C.

Author Information Supermatrix used for phylogenetic analysis deposited at Morphobank (http://www.morphobank.org/) project P888. The new taxa (new genus and new species) have been registered with Zoobank (LSID: urn:lsid:zoobank.org:pub:415D35BC-3306-4361-B34D-B7702649B425). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.G. (jgeisler@nyit.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13238

Rapid and tunable post-translational coupling of genetic circuits

Arthur Prindle[1]*, Jangir Selimkhanov[1]*, Howard Li[1], Ivan Razinkov[1], Lev S. Tsimring[2] & Jeff Hasty[1,2,3]

One promise of synthetic biology is the creation of genetic circuitry that enables the execution of logical programming in living cells. Such 'wet programming' is positioned to transform a wide and diverse swathe of biotechnology ranging from therapeutics and diagnostics to water treatment strategies. Although progress in the development of a library of genetic modules continues apace[1–4], a major challenge for their integration into larger circuits is the generation of sufficiently fast and precise communication between modules[5,6]. An attractive approach is to integrate engineered circuits with host processes that facilitate robust cellular signalling[7]. In this context, recent studies have demonstrated that bacterial protein degradation can trigger a precise response to stress by overloading a limited supply of intracellular proteases[8–10]. Here we use protease competition to engineer rapid and tunable coupling of genetic circuits across multiple spatial and temporal scales. We characterize coupling delay times that are more than an order of magnitude faster than standard transcription-factor-based coupling methods (less than 1 min compared with ~20–40 min) and demonstrate tunability through manipulation of the linker between the protein and its degradation tag. We use this mechanism as a platform to couple genetic clocks at the intracellular and colony level, then synchronize the multi-colony dynamics to reduce variability in both clocks. We show how the coupled clock network can be used to encode independent environmental inputs into a single time series output, thus enabling frequency multiplexing (information transmitted on a common channel by distinct frequencies) in a genetic circuit context. Our results establish a general framework for the rapid and tunable coupling of genetic circuits through the use of native 'queueing' processes such as competitive protein degradation.

To engineer rapid coupling between synthetic genetic modules, we developed a post-translational coupling platform that operates via shared degradation by the ClpXP protease (Fig. 1a). In this scheme, all LAA-tagged components[11] are dynamically linked through competition for a limited number of proteases[10,12], such that tagged modules remain tightly aligned (1 ± 1 min (± s.e.m.), GFP–CFP (green fluorescent protein–cyan fluorescent protein) experimental trajectory pairs in Fig. 1a) despite significant induction delay (31 ± 5 min, time elapsed from addition of inducer to GFP appearance in Fig. 1a). This coupling method produces delays that are more than an order of magnitude faster than standard transcription-factor-based coupling methods (~20–40 min)[13,14]. To illustrate directly the response time that can be achieved by coordinating module output via modulating ClpXP activity, we show that low levels (90 μM) of externally provided H_2O_2 'inducer' rapidly (<2 min, our experimental time-step) and reversibly modulates the concentration of constitutively expressed GFP in a ClpXP-dependent manner (Fig. 1b). Here, H_2O_2 reduces the native substrate load on ClpXP by obstructing RssB, the adaptor protein that targets the alternative sigma factor RpoS for degradation by ClpXP[8,9,15]. As RpoS is continuously produced and degraded by ClpXP, inactivating its rate-limiting adaptor protein results in an instantaneous increase in the effective ClpXP degradation rate for LAA-tagged proteins[16].

We systematically explored the coupling mechanism by driving a constitutive module with a quorum-sensing clock (Fig. 1c). As the pacemaker, the quorum clock generates density-dependent synchronous oscillations at the colony level via acyl-homoserine lactone (AHL), a small molecule capable of synchronizing cellular behaviour across distances up to 100 μm[17]. Using microfluidic devices[18] we observed the colony-level expression of the constitutive module, and found that oscillating expression was synchronized to the quorum clock (Fig. 1c, top right). We then constructed a library of degradation tags by adding a series of variable-length spacer regions between the downstream protein and its degradation tag. Spacer regions contained between one and five copies of the amino acid sequence 'Thr-Ser' (TS) and their effects on offset time compared to that of a previously published alternative degradation tag (Extended Data Fig. 1b–f). Although all spacer sequences produced synchronous activation dynamics, the degradation dynamics of the downstream module were offset depending on the length of the linker sequence, where longer linkers produced greater GFP–CFP offset time (Fig. 1c, bottom). Thus, our ClpXP coupling platform rapidly links genetic modules through shared degradation and permits tuning the strength and timing of coupling by changing the degradation kinetics of individual modules.

To engineer coupling between genetic modules capable of generating their own dynamics, we designed a circuit containing the quorum clock and a variant of a previously described intracellular clock (Fig. 2a)[19]. This intracellular clock variant based on the $P_{lac/ara-1}$ promoter retains the fast dynamics and simple genetic architecture of the published version that uses the $P_{LlacO-1}$ promoter, yet its period is tunable by both isopropyl-β-D-1-thiogalactopyranoside (IPTG) and arabinose in the presence of chromosomal araC. We first used small microfluidic devices (100 cells) and observed fast and asynchronous intracellular clock oscillations without quorum clock contribution, as the quorum clock requires a critical colony size to function (Supplementary Video 1). In larger devices (5,000 cells), we observed a transition from asynchronous oscillations to identical intracellular and quorum clock oscillations as the population grew larger (Fig. 2b and Supplementary Video 2). In the case of the larger population, the substrate load on ClpXP during the quorum clock pulse is sufficient to shift the intracellular clock out of its oscillatory regime, enabling complete linkage between the two clocks despite their vastly different spatial and temporal scales. Thus, despite lacking a mode of cell–cell communication itself, the intracellular clock is effectively synchronized at the colony level through ClpXP-mediated coupling with the quorum clock.

We found that changing the intracellular clock period of individual cells indirectly tuned the quorum clock period, as IPTG values associated with longer intracellular clock periods inversely produced shorter quorum clock periods (Fig. 2c). We developed a computational model of the oscillator network involving a form of load-mediated pulse frequency modulation to explain this effect (Fig. 2d–f). Between coupled pulses, the intracellular clock accelerates the quorum pulse onset through load-mediated decreases in the degradation rate of LuxI, since larger intracellular

[1]Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA. [2]BioCircuits Institute, University of California, San Diego, La Jolla, California 92093, USA. [3]Molecular Biology Section, Division of Biological Science, University of California, San Diego, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

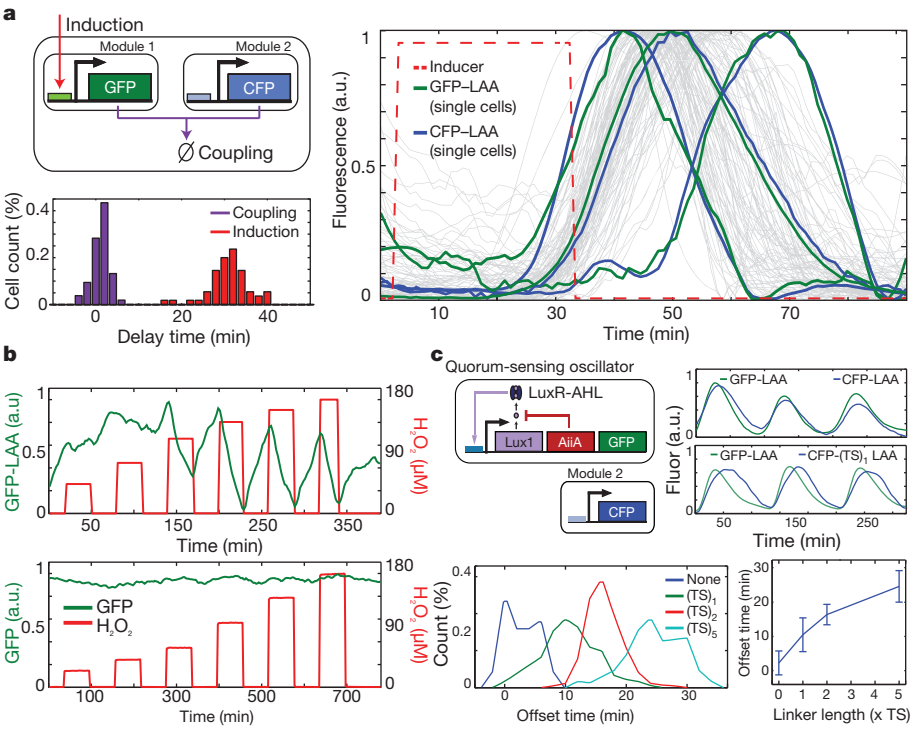


Figure 1 | A rapid post-translational coupling platform based on shared degradation. **a**, We measured the delays associated with module-module coordination by ClpXP degradation (1 ± 1 min (± s.e.m.), represented by purple arrow) and input–output response through transcription and translation (31 ± 5 min) in a single experiment by inducing (red arrow) the *lux* promoter and tracking the response of superfolder GFP (sfGFP)-LAA (*lux* promoter, black arrow) and CFP-LAA ($P_{lac/ara-1}$ promoter, black arrow) in single cells (right panel, 55 gray cell trajectory pairs with 3 representative pairs highlighted). **b**, Rapid (<2 min, our experimental time-step) induction of protein degradation by externally provided H_2O_2 produces reversible changes in ClpXP load in response to obstruction of RssB[8,9,15]. **c**, To use post-translational coupling to drive downstream modules, we linked a quorum clock to a constitutively expressed fluorescent protein via the addition of identical LAA tags. With identical degradation tags, the constitutive module couples tightly to the quorum pacemaker. The addition of a variable-length linker (Thr-Ser (TS) repeats) before the degradation tag phase-shifts the degradation dynamics, where longer linkers produced longer delays. Error bars indicate s.d. of offset time, centred at the mean (50–200 cells for each TS-linker length). a.u., arbitrary units.

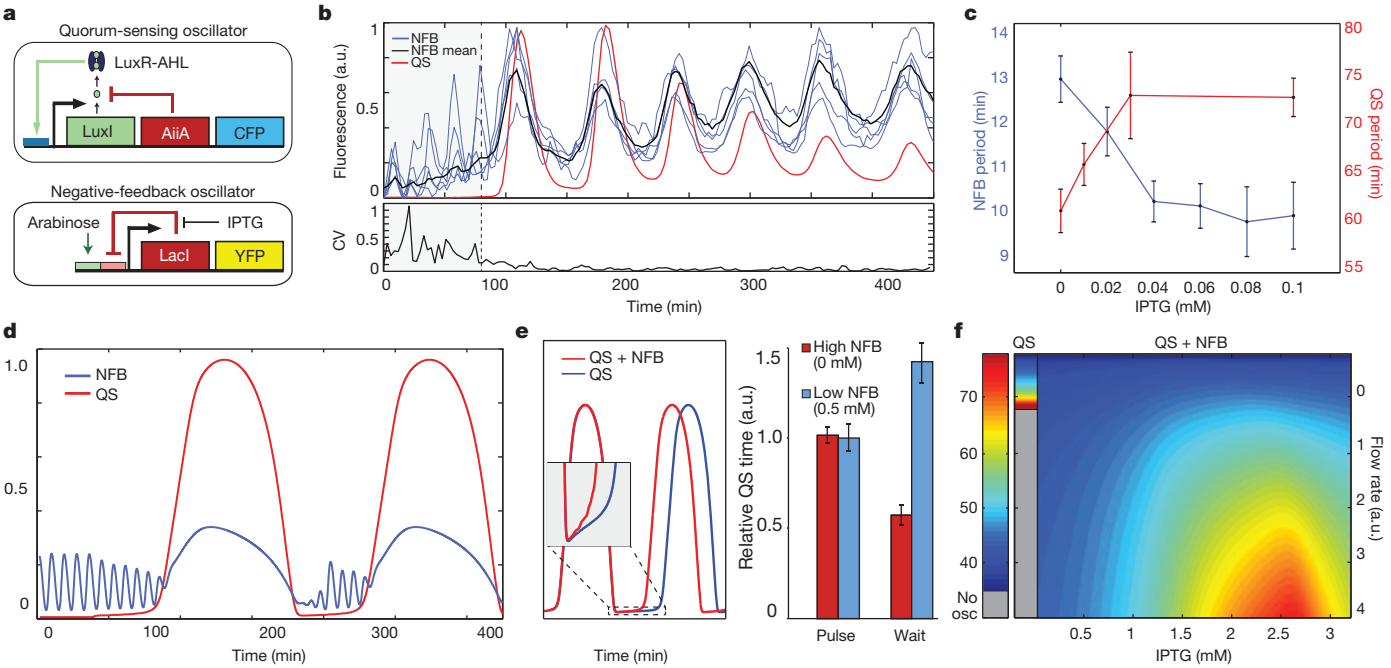


Figure 2 | Post-translationally linked genetic clocks at multiple scales. **a**, The network is composed of coupled intracellular[19] and quorum clocks[17]. The intracellular clock oscillates as a result of delayed negative feedback on its own promoter and its period is tunable by IPTG and arabinose. Quorum clock oscillations are tunable by media flow rate and are synchronized via AHL at the colony level. **b**, The coupled intracellular quorum clock system oscillates asynchronously in small populations and transitions to synchronized oscillations in larger populations once the quorum clock fires. Despite lacking a mode of cell–cell communication itself, the coefficient of variation (CV) of the intracellular clock drops markedly through host-linked coupling with the quorum clock (bottom, data from 28 single-cell traces). **c**, IPTG reduces the intracellular clock period in small cell populations without the quorum clock (blue) and increases the coupled period in larger populations with the quorum clock (red). Each data point taken from 10–30 oscillatory peaks. Error bars indicate s.e.m. of the period, centred at the mean. **d**, In our computational model, load-mediated coupling allows the intracellular clock to modulate the quorum clock period via degradation coupling at ClpXP, since the intracellular clock continues oscillating between coupled pulses and accelerates the pulse onset. **e**, This adaptive form of pulse frequency modulation ensures that the pulse dynamics remain unchanged while the inter-pulse duration is adjusted (left, model; right, experimental; 6–9 oscillatory peaks). Inset shows the earlier onset of the coupled pulse due to the intracellular clock. Error bars indicate s.e.m. of relative quorum clock period. **f**, This mechanism also makes the coupled system more robust by enabling oscillation at higher media flow rates. NFB, negative-feedback oscillator; QS, quorum-sensing oscillator.

©2014 Macmillan Publishers Limited. All rights reserved

clock load produces higher levels of the AHL-synthase (Fig. 2e, left, and Extended Data Fig. 2a–e). During the coupled pulse, contributions of the intracellular clock leave the duration of the pulse itself unchanged (Fig. 2e, left (model) and right (experimental)). Linking the intracellular and quorum clocks through degradation also yielded an expansion in the oscillatory regime for the coupled system with respect to flow rate compared to the quorum clock alone (Fig. 2f). In this way, the intracellular clock continually excites the quorum clock to fire, enabling more robust function at higher external flow rates (Extended Data Fig. 3a–c).

With a platform for rapidly coupling genetic clocks at multiple scales, we sought to engineer a system capable of frequency-encoding information from both clocks into the multispectral time series of a single reporter (Fig. 3a). Here, the measured output of the intracellular clock reporter contains contributions from its own fast intracellular clock dynamics between slow quorum clock bursts (Supplementary Video 3). As the range of natural periods for the faster $P_{lac/ara\text{-}1}$ intracellular clock is fully separated from the slower quorum clock[17,19], both IPTG and arabinose and flow rate inputs can be encoded into frequency-modulated oscillations in the time domain and independently extracted by Fourier transform. Thus, the measurement of a single clock history reveals the activities both underlying clock networks.

We began by characterizing the frequency response curves for both the intracellular and quorum clocks in isolation, finding ranges of 7–25 min and 55–95 min, respectively, when sweeping IPTG and arabinose, and flow rate inputs (Fig. 3b, top (intracellular clock in araC+ strain) and bottom (quorum clock, original study data)[17]). We then measured trajectories taken from the coupled clock system and extracted the frequency components of both clocks by Fourier transform[20] (Fig. 3c and Methods Summary). In sweeping IPTG and arabinose inducers, we found the frequency response of the intracellular clock contribution to the multispectral reporter to be unchanged by the inclusion of the quorum clock since the intracellular frequency response to IPTG and arabinose was equivalent to the isolated clock (Fig. 3d, top (coupled), and Fig. 3b, top (isolated)). We then swept flow rates at three fixed inducer levels, and found distinct response curves for the quorum clock contribution to the multispectral reporter shifted in accordance with our model for ClpXP-mediated frequency modulation by the intracellular clock (Fig. 3d, bottom). Thus, to decode a given pair of IPTG and arabinose, and flow rate inputs, we first recover the intracellular clock frequency as a measure of IPTG and arabinose and then use the corresponding quorum clock response curve to measure flow rate.

To extend rapid coupling to greater spatial scales, we added a genetic H_2O_2 signalling[21] cassette to the network and observed synchronization at the multi-colony level (Fig. 4a and Supplementary Video 4). In conducting these experiments, we also observed H_2O_2-mediated interaction between the native stress response network and our synthetic circuit at ClpXP (Fig. 4b). In the original design, H_2O_2 synchronized quorum clock oscillations by transcriptional upregulation of the *lux* promoter via the aerobic response control system ArcAB[21]. In addition to transcriptional increase (Fig. 4c, top), we found an increase in the apparent degradation rate with H_2O_2 (Fig. 4c, bottom, and Extended Data Fig. 4a, b), consistent with increased ClpXP activity in response to externally provided H_2O_2. The coupled increases in transcriptional output and effective ClpXP degradation rate in response to H_2O_2 also tightens the period distribution at the multi-colony level by mitigating the effects of period variation in an individual colony (Fig. 4c, top, and Extended Data Fig. 4c, d).

Engineering synthetic circuits composed of interacting modules is an ongoing effort[1–6] that has generally relied on transcription and translation, with less attention paid to post-translational coupling mechanisms[22].

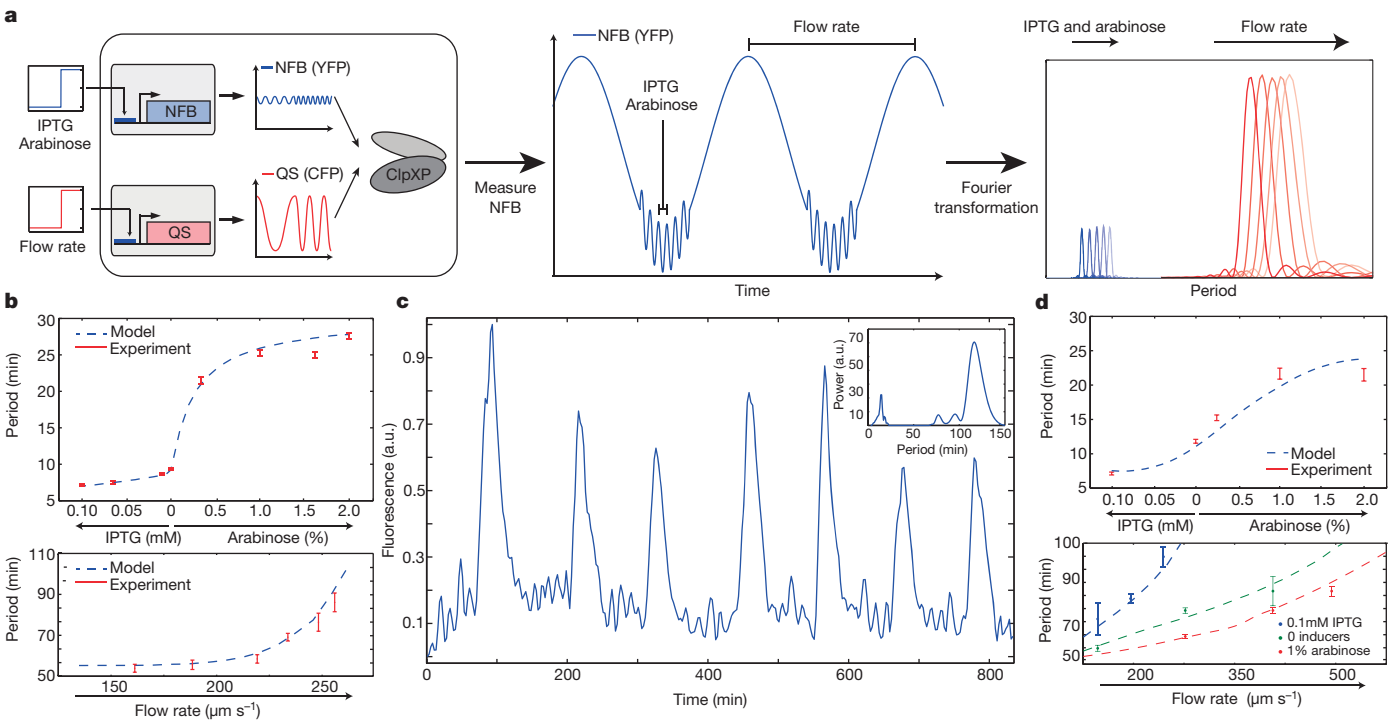


Figure 3 | Genetic multispectral encoding. **a**, Separate IPTG and arabinose, and flow rate inputs are encoded into frequency-modulation oscillations that can be measured from the time series of the reporter for the intracellular clock. This engineered system is capable of encoding information from two underlying networks into a single multispectral time series. **b**, Frequency response curves generated from experimental data and computational models for the intracellular clock (top, data from 30 single-cell traces each) and quorum clock (bottom, model applied to data from the original study[17]) in isolation. Error bars indicate s.e.m. of the period, centred at the mean. **c**, In the coupled system, frequency-modulated oscillations from both clocks can be observed in the output of the intracellular clock and extracted by inverse Fourier transformation (inset, methods in Supplementary Information. **d**, Independent recovery of both IPTG and arabinose and flow rate inputs. The frequency response of the intracellular clock to IPTG and arabinose is equivalent to the isolated clock (top) and the frequency response of the quorum clock is shifted by the intracellular clock (bottom). Periods calculated from 5–10 single-cell traces for each condition. Error bars indicate s.e.m. of the period, centred at the mean.

©2014 Macmillan Publishers Limited. All rights reserved

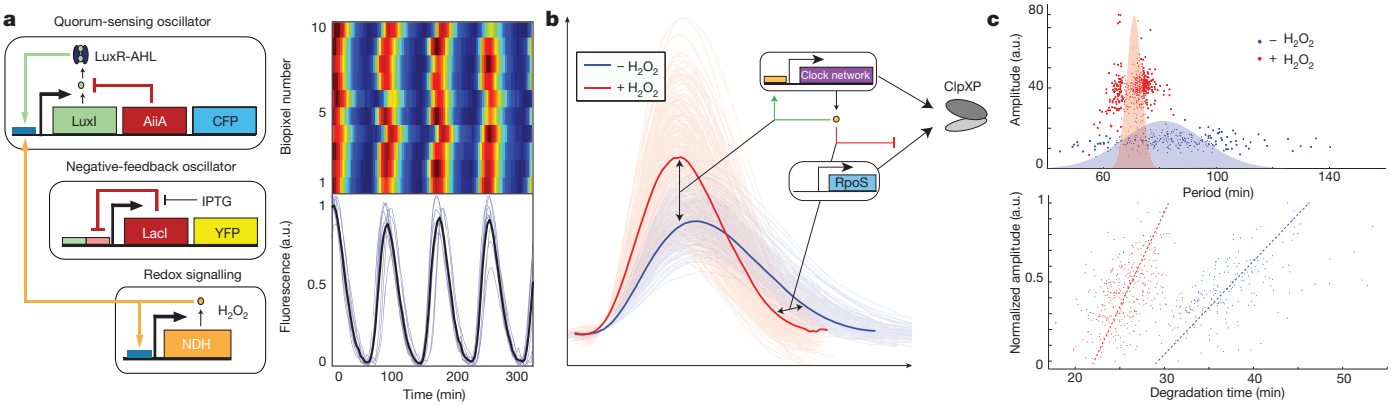


Figure 4 | Post-translational coupling at the multi-colony level. **a**, At the multi-colony level, interaction of H_2O_2 generated by redox signalling with the cellular stress response network synchronizes quorum clock oscillations between colonies. Traces taken from 10 separate colonies across the array. **b**, Host-linked oscillations change distinct aspects of the waveform in response to H_2O_2 produced by the enzymatic activity of NDH (NADH dehydrogenase II). With H_2O_2, oscillations have larger amplitudes and steeper downslopes, revealing increases in both transcription and degradation produced by the interaction of the synthetic clock network with the native stress response. Dark lines indicate the means of all trajectories. **c**, H_2O_2 increases the oscillatory amplitude while decreasing the required degradation time, revealing an increase in ClpXP activity. This increase in ClpXP capacity in response to H_2O_2 serves to mitigate the effects of transcriptional noise by minimizing the effects of amplitude variation on the period, resulting in a tightening of the period distribution with H_2O_2 (model, Extended Data Fig. 4c, d).

Protease competition offers the advantages of rapid response, modularity with distinct recognition sequences, and simultaneous control over multiple circuits with protease adapters[23,24]. More generally, in natural biological networks, competition for cellular resources (for example, metabolites, enzymes, transcription factors, binding sites) produces nonlinear coupling effects that serve to reduce noise, increase sensitivity to input concentrations, and discriminate between multiple inputs[12,25–28]. We envision that coordinating engineered circuits via built-in cellular processes—what we term 'host-linked' coupling—has the potential to produce more sophisticated circuits by facilitating robust signalling between synthetic modules.

METHODS SUMMARY

Strains and plasmids. The oscillator plasmids were constructed by modifying and combining published constructs[17,19,21] by PCR reactions and all circuit components except *luxR* were tagged by PCR with a carboxy-terminal ssrA tag (AANDENYA LAA)[11] for fast degradation. We placed the activator and reporting elements (LuxI, CFP, and YFP) on one vector (IRAP2, Kan and ColE1) and the repressing elements (AiiA and LacI) on a second vector (IRAP3, Amp and p15A). The Thr-Ser (TS) constructs were constructed by adding various TS repeat inserts between the CFP and the LAA tag. For example, for two TS, the amino acid sequence 'TSTS' was inserted immediately before the degradation tag 'AANDENYALAA'. The AAV construct was constructed by replacing the 'LAA' portion of the degradation tag with 'AAV'.

Microfluidics and microscopy. Image acquisition was performed on a Nikon TI and images were acquired using a Photometrics CoolSnap cooled CCD camera or Photometrics QuantEM EMCCD camera, both controlled by Nikon Elements software. The cells were imaged inside a microfluidic device with the ability to mix or switch between two different media sources. On the day of the experiment, 50 µl of an overnight culture was diluted in 50 ml of lysogeny broth (Difco) and antibiotics. When cells reached a $D_{600\ nm}$ of 0.1, cells were spun down and resuspended in 5 ml of fresh media and loaded into the device. Three devices were used to study populations of varying sizes: small colony (100 cells)[29], large colony (5,000 cells)[17], and multiple large colonies (500 colonies of 5,000 cells)[21].

Data analysis. Single cell and individual trap fluorescent trajectories were obtained from time-lapse images using our previously developed algorithms[21,29] and built-in MATLAB functions. We identified peaks and troughs from these trajectories and used these values to calculate periods and amplitudes. To calculate the coupling delay in Fig. 1a and offset time in Fig. 1c, we measured the difference between the 10% amplitude points of trajectory pairs. The induction time was measured from induction start time to 10% amplitude of the induced module. To extract both frequencies from time series data, we performed Fourier transformations using the Lomb–Scargle algorithm. We used two sequential transformations to isolate each component separately. First, we used a band-pass filter (5–25 min) to extract the fast intracellular clock component. Then, we filtered out these fast frequencies using a second band-pass filter (75–150 min) to extract the slower quorum clock component. Finally, we overlay the two power spectra, preserving the relative amplitude of the peaks.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 June 2013; accepted 10 March 2014.
Published online 9 April 2014.

1. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch from *Escherichia coli*. *Nature* **403,** 339–342
2. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnol.* **31,** 448–452 (2013).
3. Tigges, M., Marquez-Lago, T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457,** 309–312 (2009).
4. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333,** 1307–1311 (2011).
5. Moon, T. S., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491,** 249–253 (2012).
6. Del Vecchio, D., Ninfa, A. J. & Sontag, E. D. Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.* **4,** 161 (2008).
7. Nandagopal, N. & Elowitz, M. B. Synthetic biology: integrated gene circuits. *Science* **333,** 1244–1248 (2011).
8. Fredriksson, Å. *et al.* Decline in ribosomal fidelity contributes to the accumulation and stabilization of the master stress response regulator σs upon carbon starvation. *Genes Dev.* **21,** 862–874 (2007).
9. Merrikh, H., Ferrazzoli, A. E., Bougdour, A., Olivier-Mason, A. & Lovett, S. T. A DNA damage response in *Escherichia coli* involving the alternative sigma factor, RpoS. *Proc. Natl Acad. Sci. USA* **106,** 611–616 (2009).
10. Cookson, N. A. *et al.* Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7,** 561 (2011).
11. Keiler, K. C., Waller, P. & Sauer, R. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271,** 990–993 (1996).
12. Goldbeter, A. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78,** 6840–6844 (1981).
13. Rosenfeld, N. & Alon, U. Response delays and the structure of transcription networks. *J. Mol. Biol.* **329,** 645–654 (2003).
14. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102,** 3581–3586 (2005).
15. Mika, F. & Hengge, R. A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of σs (RpoS) in *E. coli*. *Genes Dev.* **19,** 2770–2781 (2005).
16. Pruteanu, M. & Hengge-Aronis, R. The cellular level of the recognition factor rssb is rate-limiting for σs proteolysis: implications for rssb regulation and signal transduction in σs turnover in *Escherichia coli*. *Mol. Microbiol.* **45,** 1701–1713 (2002).
17. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463,** 326–330 (2010).
18. Unger, M. A., Chou, H.-P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science,* **288,** 113–116 (2000).
19. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456,** 516–519 (2008).

©2014 Macmillan Publishers Limited. All rights reserved

20. Press, W. H. in *Numerical Recipes: The Art of Scientific Computing* 3rd edn (Cambridge Univ. Press, 2007).
21. Prindle, A. *et al.* A sensing array of radically coupled genetic 'biopixels'. *Nature* **481,** 39–44 (2012).
22. Grünberg, R. & Serrano, L. Strategies for protein synthetic biology. *Nucleic Acids Res.* **38,** 2663–2675 (2010).
23. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering controllable protein degradation. *Mol. Cell* **22,** 701–707 (2006).
24. Griffith, K. L. & Grossman, A. D. Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Mol. Microbiol.* **70,** 1012–1025 (2008).
25. Burger, A., Walczak, A. M. & Wolynes, P. G. Abduction and asylum in the lives of transcription factors. *Proc. Natl Acad. Sci. USA* **107,** 4016–4021 (2010).
26. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature Genet.* **43,** 854–859 (2011).
27. Buchler, N. E. & Louis, M. Molecular titration and ultrasensitivity in regulatory networks. *J. Mol. Biol.* **384,** 1106–1119 (2008).
28. Strogatz, S. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry and Engineering* (Perseus Books, 2001).
29. Ferry, M. S., Razinkov, I. & Hasty, J. Microfluidics for synthetic biology from design to execution. *Methods Enzymol.* **497,** 295–372
30. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64,** 2240–2246 (1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Science Foundation (MCB-1121748) and by the San Diego Center for Systems Biology (NIH Grant P50 GM085764) and the US Department of Defense National Defense Science and Engineering Graduate Fellowship (A.P.). We would like to thank T. Danino, M. Jin, C. Rivera, O. Din and J. De Friel for critical reading of the manuscript.

Author Contributions All authors (A.P., J.S., H.L., I.R., L.T., and J.H.) contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H. (hasty@bioeng.ucsd.edu).

©2014 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature13070

Apical constriction drives tissue-scale hydrodynamic flow to mediate cell elongation

Bing He[1]*, Konstantin Doubrovinski[1]*, Oleg Polyakov[2] & Eric Wieschaus[1,3]

Epithelial folding mediated by apical constriction converts flat epithelial sheets into multilayered, complex tissue structures and is used throughout development in most animals[1]. Little is known, however, about how forces produced near the apical surface of the tissue are transmitted within individual cells to generate the global changes in cell shape that characterize tissue deformation. Here we apply particle tracking velocimetry in gastrulating *Drosophila* embryos to measure the movement of cytoplasm and plasma membrane during ventral furrow formation[2,3]. We find that cytoplasmic redistribution during the lengthening phase of ventral furrow formation can be precisely described by viscous flows that quantitatively match the predictions of hydrodynamics. Cell membranes move with the ambient cytoplasm, with little resistance to, or driving force on, the flow. Strikingly, apical constriction produces similar flow patterns in mutant embryos that fail to form cells before gastrulation ('acellular' embryos), such that the global redistribution of cytoplasm mirrors the summed redistribution occurring in individual cells of wild-type embryos. Our results indicate that during the lengthening phase of ventral furrow formation, hydrodynamic behaviour of the cytoplasm provides the predominant mechanism transmitting apically generated forces deep into the tissue and that cell individualization is dispensable.

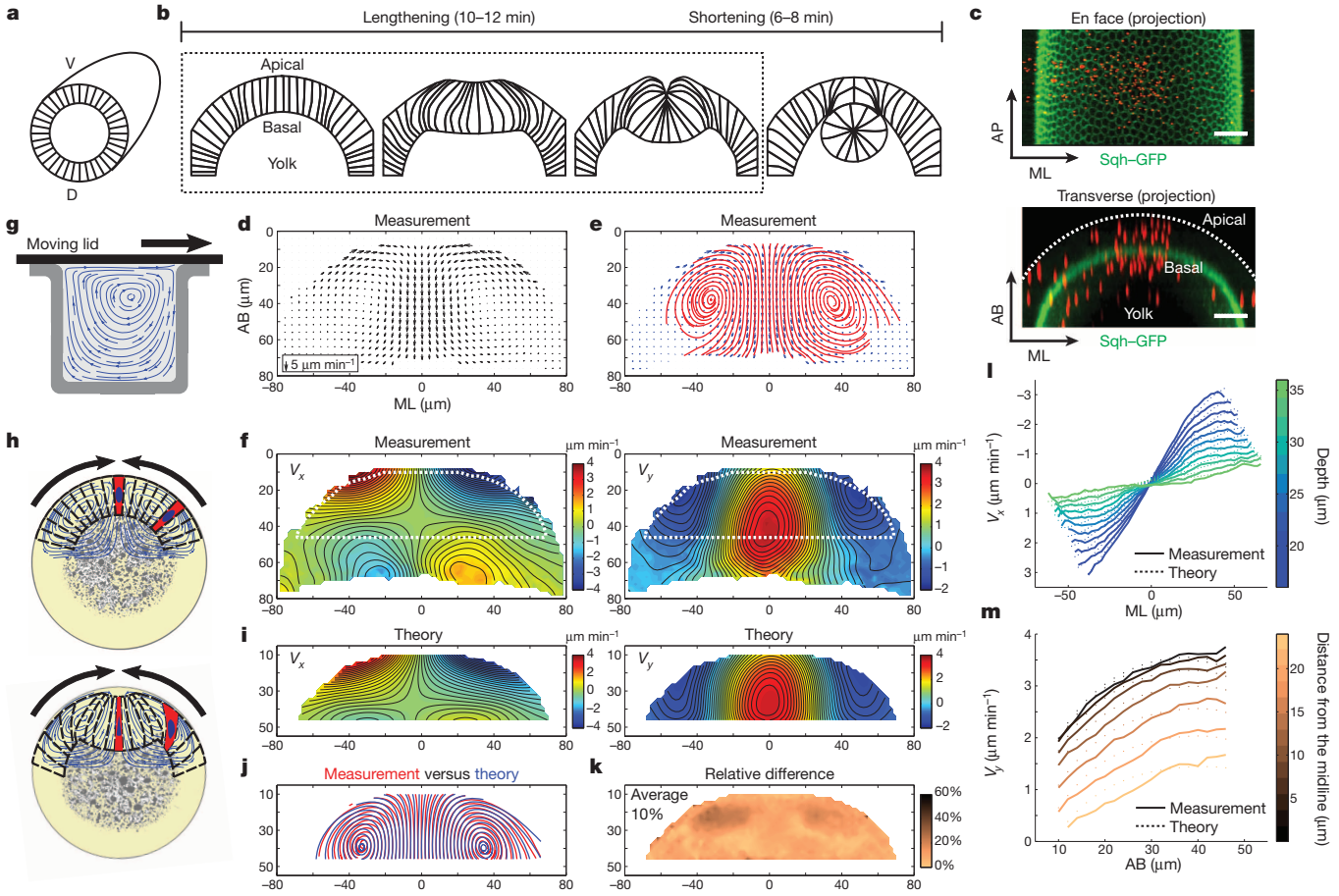


Figure 1 | Cytoplasmic flow during ventral furrow formation. **a**, **b**, Cross-section view of ventral furrow formation. V, ventral; D, dorsal. **c**, An embryo injected with fluorescent beads (red). Scale bar, 20 μm. AP, anterior–posterior; ML, medial–lateral; AB, apical–basal. **d**, **e**, The velocity field (arrows) and streamlines (red) of the cytoplasmic flow at $t = 4$–6 min ($n = 14$ embryos). **f**, Heat maps of V_x and V_y with smoothed contour lines of equal magnitude. Positive values indicate left-to-right flow (V_x) or basally directed flow (V_y). The dotted line highlights the region subjected to theoretical comparison. **g**, A two-dimensional Stokes flow driven by a moving lid. **h**, Apical constriction drives cytoplasmic flow to mediate cell shape changes. **i**, V_x and V_y deduced from the Stokes equations. **j**, Streamlines of the measured (red) and deduced (blue) velocity fields. **k**, Relative difference in **j**. **l**, **m**, V_x (**l**) and V_y (**m**) as a function of ML or AB positions, respectively.

[1]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [2]Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. [3]Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

During *Drosophila* gastrulation, ventrally localized prospective mesoderm forms a furrow and invaginates from the surface of the embryo. As this furrow forms, individual cells first constrict at their apical end and undergo elongation ('cell lengthening'). During their subsequent invagination, cells shorten back to wedge-like shapes (Fig. 1a, b). Apical constriction is powered by an apically localized contractile actomyosin network that forms at the onset of gastrulation and is widely believed to be the main force driving ventral furrow formation[4]. Although it is unclear how stresses transmitted from the apical cortex mediate tissue movement in the interior of the embryo, a common view is that the cell surface and associated cytoskeleton play a major role in generating and transmitting forces within each cell, while the cytoplasm passively adopts the shape defined by the cortex. In an opposing view, the entire tissue is considered as a continuum; forces are transmitted continuously across the epithelium and its subdivision into cells is not a fundamental component in force transmission. Modelling studies have successfully described global tissue movements using both cell-based[5–7] and continuum viewpoints[8–11]. Although these studies identified plausible mechanisms that could mediate ventral furrow morphogenesis, the actual mechanism remains elusive. One fundamental limitation is that previous measurements of tissue deformation have an intrinsically limited spatial resolution of a single cell. A rigorous test of physical mechanisms, however, requires tissue deformation being tracked with subcellular resolution[12].

With this goal in mind, we developed a strategy using injected submicrometre fluorescent beads as passive tracers to measure the motion of cytoplasm in *Drosophila* embryos[13,14] (Fig. 1c and Extended Data Fig. 1). The injected beads show extremely low mobility when there is no global movement of the tissue (Extended Data Fig. 2 and Supplementary Methods), thereby providing trackable landmarks for quantitative, high-resolution measurements of cytoplasmic flows (Extended Data Figs 3 and 4 and Supplementary Video 1). Figure 1d–f shows averaged two-dimensional velocity distribution and streamlines at a transverse cross section through the embryo at the middle of the lengthening phase ($t = 4$–6 min; Supplementary Methods). The flow patterns do not change substantially during the lengthening phase (Extended Data Fig. 4a). When the shortening phase starts, the tracking of beads becomes difficult as beads locate deeper into the embryo. We therefore focused our analysis on the lengthening phase.

The measured movement of cytoplasm resembles a laminar flow (Fig. 1g) in that when the cortex constricts and moves, the underlying cytoplasm seems to be dragged with it (Fig. 1h). If this is in fact the case, tissue deformation in the interior of the embryo should follow the Stokes equations that describe the dynamics of viscous flows in the low Reynolds number regime. In particular, flow velocity at any point in the interior of an arbitrarily chosen domain will be uniquely determined by the velocity distribution at the domain boundaries. This allows for a quantitative and parameter-free comparison between Stokes dynamics and tissue movements (Methods). We found that the inferred velocity distribution is in close agreement with our measurements, with relative differences close to the measurement error (~10%; Fig. 1i–m and Extended Data Fig. 4c). The close agreement strongly argues that tissue deformation during the lengthening phase could arise exclusively from the viscous response of the cytoplasm to the shear force generated by apical constriction at the cortex (Supplementary Notes).

This close agreement implies that lateral membranes do not exert appreciable forces on the cytoplasm and thus are expected to (1) co-move with the cytoplasmic flow and (2) be dispensable for the redistribution of the cytoplasm that underlies cell shape changes. We tested prediction 1 by tracking the movement of wheat germ agglutinin (WGA)-coated beads attached to the plasma membrane[15] (Fig. 2a, Extended Data Fig. 5 and Supplementary Videos 2 and 3). During gastrulation, the motion of such beads is very similar to non-WGA cytoplasmic beads, with a relative difference of only 20% at $t = 5$ min (Fig. 2b–d, Extended Data Fig. 6 and Supplementary Methods). Moreover, the flow of the WGA beads also matches the predictions of the Stokes equations with 86% similarity (Fig. 2f–i). Most interestingly, the expansion of the lateral

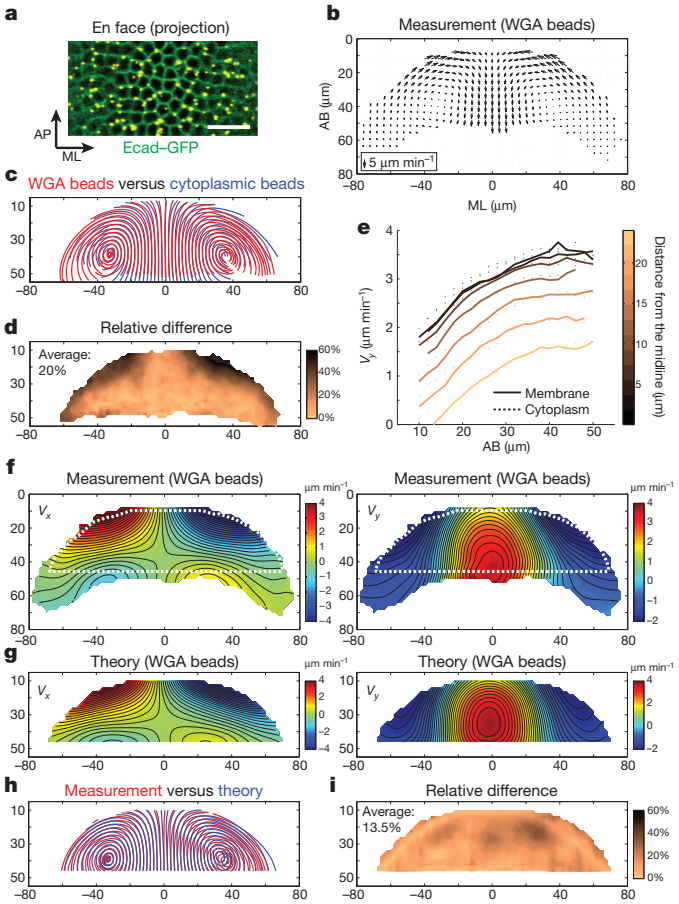


Figure 2 | The movement and expansion of the lateral membranes follow the cytoplasmic flow. **a**, Perivitelline injection of WGA-coated beads (yellow). Scale bar, 20 μm. **b**, Velocity field of the WGA beads at $t = 4$–6 min ($n = 10$ embryos). **c**, Streamlines of the WGA beads (red) and the cytoplasmic beads (blue). **d**, Relative difference in **c**. **e**, V_y as a function of AB positions. **f**, **g**, Heat maps of V_x and V_y of the WGA beads: **f**, measurement; **g**, theoretical prediction. **h**, Streamlines of the measured (red) and deduced (blue) velocity fields. **i**, Relative difference in **h**.

membranes closely matches the regional flow of the ambient cytoplasm, being highest apically and decreasing towards the base (Figs 1m and 2e). These data are very consistent with a model where lateral membranes extend passively as a consequence of the cytoplasmic flow without offering appreciable driving force or resistance.

To test prediction 2, we took advantage of the fact that *Drosophila* begins its development as a syncytium and only forms individual cells with basolateral membranes immediately preceding gastrulation. We have found that simultaneous elimination of two zygotically active genes (*CG9506* (=*slam*[16]) and *CG34137*) blocks the formation of basolateral membranes, while maintaining the normal subcellular organization of the cytoplasm (Fig. 3a, Extended Data Fig. 7 and Supplementary Videos 4, 5 and 7). In such embryos, the expression pattern of the mesoderm determinants Twist and Snail[17] is normal (Fig. 3c and Extended Data Fig. 8a), and gastrulation starts at the normal time, with myosin forming a cortical network that undergoes dynamic, pulsed contractions on the ventral surface (Supplementary Video 6 and Extended Data Fig. 8b, c). The average rate of the resultant apical constriction, however, is reduced to 60% of that in the wild type (Extended Data Fig. 9), presumably because the network is less well organized without cell individualization. As in wild-type embryos, apical constriction leads to formation of membrane blebs, albeit larger, on the ventral surface of the gastrulating acellular embryos[3] (Extended Data Fig. 8d).

Remarkably, apical constriction in the acellular embryos leads to basal movement of bulk cytoplasm and nuclei towards the yolk in a manner

©2014 Macmillan Publishers Limited. All rights reserved

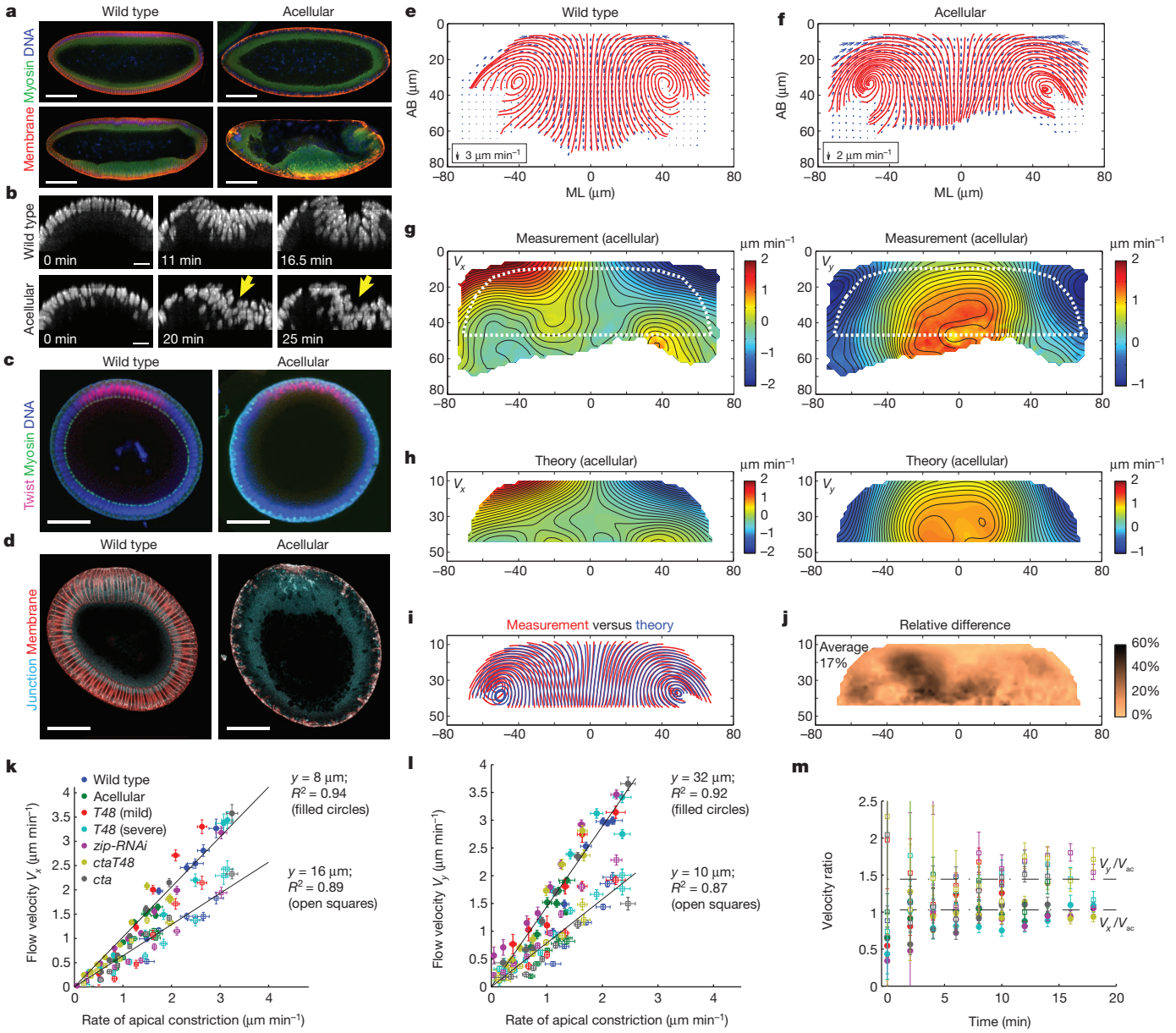


Figure 3 | Apical constriction induces cytoplasmic flow independent of the basolateral membranes. **a**, Midsagittal view of embryos showing membrane, myosin II and DNA. Top, cellularization; bottom, early gastrulation. Scale bars, 100 μm. **b**, The nuclear movements during ventral furrow (arrows) formation. Scale bars, 20 μm. **c**, Cross section of the embryos showing Twist, myosin II and DNA. Scale bars, 50 μm. **d**, Cross section of early gastrulae showing membrane and adherens junctions. Scale bars, 50 μm. **e**, **f**, The velocity field and streamlines of the cytoplasmic flow. Wild type, $n = 20$ embryos; acellular, $n = 18$ embryos. **g**, **h**, Heat maps of V_x and V_y in the acellular embryos: **g**, measurement; **h**, theoretical prediction. **i**, Streamlines of the measured (red) and deduced (blue) velocity fields. **j**, Relative difference in **i**. **k**, V_x as a function of the rate of apical constriction (V_{ac}). **l**, Average V_y as a function of average V_{ac}. **m**, V_x/V_{ac} (filled circles) and V_y/V_{ac} (open squares) over time. Colour-coding is identical in **k**–**m**. Dashed lines, the best fit of the ratio. Error bars, s.e.m. in **k**–**m**.

similar to the wild type (Fig. 3b, d and Supplementary Video 7). The pattern of the cytoplasmic flow closely resembles that in the wild type, albeit with reduced velocity (Fig. 3e, f and Supplementary Video 8). Crucially, the reduced flow velocity quantitatively corresponds to the reduced rate of apical constriction (Extended Data Fig. 9i–k). Therefore, removing the basolateral membranes does not affect the physical response of the interior tissue to apical constriction, in strong confirmation of our model. As an extra control, we analysed four mutants that specifically affect the rate of apical constriction in otherwise normally cellularized embryos and found that in all cases the flows were well predicted by a model where apical constriction drives hydrodynamic flow of the cytoplasm (Fig. 3g–l, Extended Data Figs 9 and 10 and Supplementary Methods).

To compare cell shape changes in the wild type and acellular embryos directly, we plotted virtual cells onto the flow field and tracked their motion over time (Fig. 4a, b and Supplementary Video 9). Remarkably, virtual cells in the acellular embryos undergo morphological changes similar to those in the wild type as long as the reduced rate of apical constriction is compensated for (Fig. 4b–f and Supplementary Video 9). In particular, in both cases, virtual cell lengthening is achieved by a quasi-linear uniform extension of the apical portion of the cell (0–20 μm) (Fig. 4g). Thus, the region-specific changes in cell shape that normally occur during lengthening can be produced by laminar flow of the cytoplasm independent of the mechanical inputs from the basolateral membranes (Fig. 4h). On the other hand, the virtual cells in the acellular embryo do not undergo shortening or basal widening, and the furrow

©2014 Macmillan Publishers Limited. All rights reserved

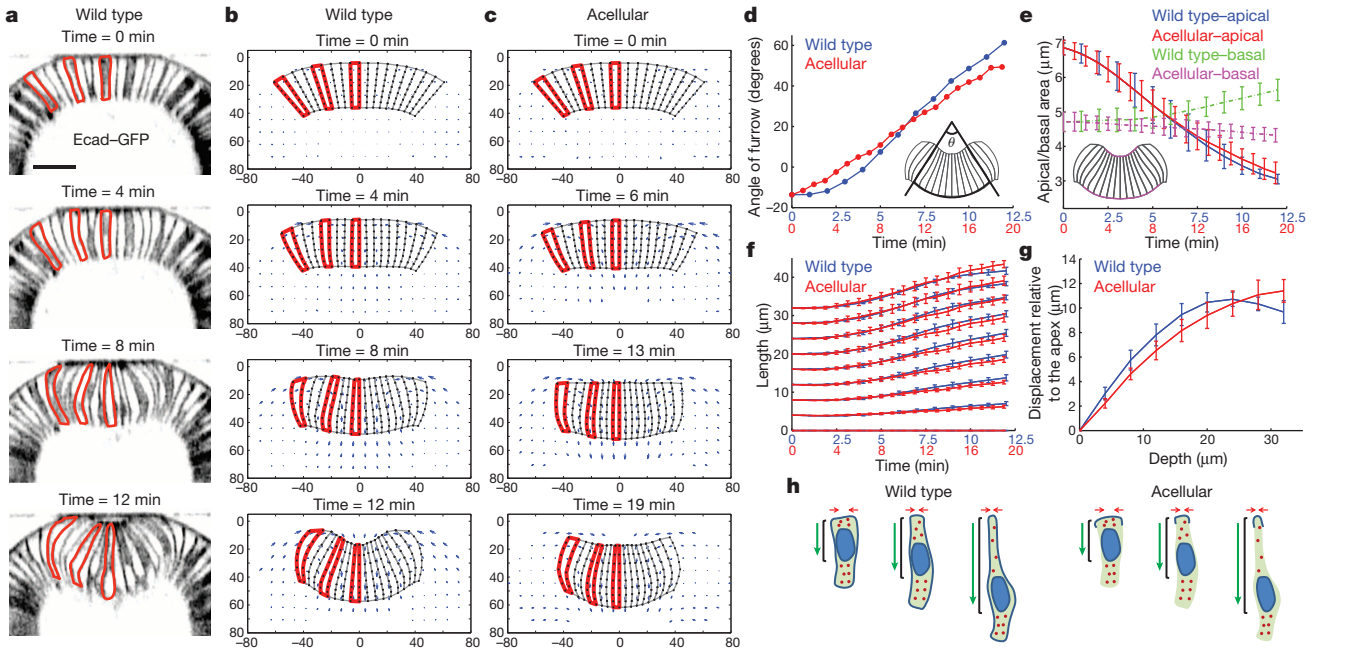


Figure 4 | Virtual cell analysis to show cell shape changes from the flow. **a**, Ventral furrow formation in the wild-type embryo. Selected cells are highlighted for better comparison. Scale bar, 30 μm. **b**, **c**, Virtual cells for the wild-type (**b**) and acellular (**c**) embryos. **d**–**f**, The secant angle enclosed by the middle 12 ventral cells (**d**), their average apical and basal area (**e**) and the average distance between the cell apex and each lateral node (**f**) as a function of time. **g**, The displacement of lateral nodes along the AB axis as a function of their initial AB positions. Error bars, s.d. in **e**–**g**. **h**, Cartoon model demonstrating that in response to apical constriction, the apical cytoplasm undergoes uniform extension independent of the basolateral membrane.

is never fully invaginated (Fig. 4e and Supplementary Videos 7–9), suggesting that some extra, cell-dependent mechanisms are required during the shortening phase.

In summary, we demonstrate that during the lengthening phase of ventral furrow formation, the behaviour of the bulk tissue below the apical cortex closely corresponds to that of a viscous fluid, although it is composed of heterogeneous cellular organelles and is partitioned into individual cells with plasma membrane boundaries. On the basis of preliminary data in our laboratory, the tissue interior is unlikely to be elastic (Extended Data Fig. 4d, Supplementary Video 10 and Supplementary Notes). In line with that, the measured flow profiles closely agree with the hydrodynamic predictions throughout the lengthening phase even after substantial tissue deformation has occurred (Extended Data Fig. 4b, c). The rate of the cytoplasmic flow is always proportional to the rate of apical constriction independent of time (Fig. 3k–m). Consistently, we found that the viscous model fits our data much better than a linear elastostatic one (90% versus 75% agreement; Supplementary Notes).

The importance of the hydrodynamic properties of the cytoplasm has been implicated in processes occurred in continuous cytoplasm, such as cytoplasmic streaming in single large cells of algae[18], the one-cell stage of *Caenorhabditis elegans* embryos[19] or the *Drosophila* oocyte[20]. Our work demonstrates that even in the context of multicellular tissues, stresses generated at the surface of the tissue can integrate with the hydrodynamic properties of the interior to transmit force and determine the specific changes in cell shape that characterize morphogenesis. This mechanism is surprisingly independent of the plasma membrane between neighbouring cells and may not require specific molecular components. Because apical constriction-induced epithelial folding occurs frequently in development (for example, *Drosophila* tracheal pit invagination[21], *Xenopus* bottle cell formation[22] and neural tube closure[23]), using viscous flow to transmit force may represent a fundamental mechanism in morphogenesis.

METHODS SUMMARY

To measure the motion of cytoplasm, we injected 500-nm red fluorescent carboxylated polystyrene microspheres (Invitrogen) into developing embryos at the late syncytial or cellularization stage using a FemtoJet express microinjector (Eppendorf). To label the plasma membrane, we coated the beads with WGA-Alexa Fluor 488 (Invitrogen) using a Carbodiimide Kit for covalent coupling (Polysciences) and injected them into the perivitelline space of the cellularizing embryos. The WGA beads bind to the plasma membrane immediately after injection and remain bound throughout development.

Injected embryos were subjected to two-photon live imaging with an excitation wavelength of 920 nm. Stacks of 40 images taken at 2-μm steps were recorded every 8 s. The images were 256 pixels × 128 pixels, corresponding to150-μm (medial–lateral) × 75-μm (anterior–posterior) regions approximately centred at the ventral midline. Using MATLAB (MathWorks) we developed methods to determine the three-dimensional trajectory of beads and to generate the velocity field of cytoplasmic flow in the course of ventral furrow formation. Each velocity field (80 × 40 grid of 2-μm squares projected on the plane of transverse cross section) was calculated by averaging bead measurements over a 2-min time span with a sampling radius of 18 μm.

The measurements and the prediction of the Stokes equations were compared as follows. We considered a sub-domain of the embryo cross section where velocity distribution could be estimated with sufficient accuracy. The boundary of this domain was specified by a discrete curve (polygon) discretized by 200 equally spaced boundary points. Stokes flow in the interior of this polygon was constructed as superposition of two-dimensional Stokeslets centred at those boundary points. The procedure allowed a parameter-free comparison between the measurements and hydrodynamic predictions.

Detailed information about reagents and methods used in this paper is described in Methods and Supplementary Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 July 2013; accepted 20 January 2014.
Published online 2 March 2014.

1. Sawyer, J. M. *et al.* Apical constriction: a cell shape change that can drive morphogenesis. *Dev. Biol.* **341,** 5–19 (2010).
2. Leptin, M. & Grunewald, B. Cell shape changes during gastrulation in *Drosophila*. *Development* **110,** 73–84 (1990).
3. Sweeton, D., Parks, S., Costa, M. & Wieschaus, E. Gastrulation in *Drosophila*: the formation of the ventral furrow and posterior midgut invaginations. *Development* **112,** 775–789 (1991).
4. Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457,** 495–499 (2009).

©2014 Macmillan Publishers Limited. All rights reserved

5. Odell, G. M., Oster, G., Alberch, P. & Burnside, B. The mechanical basis of morphogenesis. I. Epithelial folding and invagination. *Dev. Biol.* **85,** 446–462 (1981).
6. Brodland, G. W., Viens, D. & Veldhuis, J. H. A new cell-based FE model for the mechanics of embryonic epithelia. *Comput. Methods Biomech. Biomed. Eng.* **10,** 121–128 (2007).
7. Conte, V., Munoz, J. J., Baum, B. & Miodownik, M. Robust mechanisms of ventral furrow invagination require the combination of cellular shape changes. *Phys. Biol.* **6,** 016010 (2009).
8. Brodland, G. W. & Clausi, D. A. Embryonic tissue morphogenesis modeled by FEM. *J. Biomech. Eng.* **116,** 146–155 (1994).
9. Pouille, P. A. & Farge, E. Hydrodynamic simulation of multicellular embryo invagination. *Phys. Biol.* **5,** 015005 (2008).
10. Conte, V. *et al.* A biomechanical analysis of ventral furrow formation in the *Drosophila melanogaster* embryo. *PLoS ONE* **7,** e34473 (2012).
11. Kaplan, D. R. & Hagemann, W. The relationship of cell and organism in vascular plants. *Bioscience* **41,** 693–703 (1991).
12. Davidson, L. A. Embryo mechanics: balancing force production with elastic resistance during morphogenesis. *Curr. Top. Dev. Biol.* **95,** 215–241 (2011).
13. Mason, T. G., Ganesan, K., van Zanten, J. H., Wirtz, D. & Kuo, S. C. Particle tracking microrheology of complex fluids. *Phys. Rev. Lett.* **79,** 3282–3285 (1997).
14. Wirtz, D. Particle-tracking microrheology of living cells: principles and applications. *Annu Rev Biophys* **38,** 301–326 (2009).
15. Lecuit, T. & Wieschaus, E. Polarized insertion of new membrane from a cytoplasmic reservoir during cleavage of the *Drosophila* embryo. *J. Cell Biol.* **150,** 849–860 (2000).
16. Lecuit, T., Samanta, R. & Wieschaus, E. *slam* encodes a developmental regulator of polarized membrane growth during cleavage of the *Drosophila* embryo. *Dev. Cell* **2,** 425–436 (2002).
17. Leptin, M. *twist* and *snail* as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev.* **5,** 1568–1576 (1991).
18. Goldstein, R. E., Tuval, I. & van de Meent, J. W. Microfluidics of cytoplasmic streaming and its implications for intracellular transport. *Proc. Natl Acad. Sci. USA* **105,** 3663–3667 (2008).
19. Niwayama, R., Shinohara, K. & Kimura, A. Hydrodynamic property of the cytoplasm is sufficient to mediate cytoplasmic streaming in the *Caenorhabditis elegans* embryo. *Proc. Natl Acad. Sci. USA* **108,** 11900–11905 (2011).
20. Glotzer, J. B., Saffrich, R., Glotzer, M. & Ephrussi, A. Cytoplasmic flows localize injected *oskar* RNA in *Drosophila* oocytes. *Curr. Biol.* **7,** 326–337 (1997).
21. Brodu, V. & Casanova, J. The RhoGAP *crossveinless-c* links *trachealess* and EGFR signaling to cell shape remodeling in *Drosophila* tracheal invagination. *Genes Dev.* **20,** 1817–1828 (2006).
22. Hardin, J. & Keller, R. The behaviour and function of bottle cells during gastrulation of *Xenopus laevis*. *Development* **103,** 211–230 (1988).
23. Sadler, T. W. Embryology of neural tube development. *Am. J. Med. Genet. C. Semin. Med. Genet.* **135C,** 2–8 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Thiberge (Imaging core facility) for two-photon microscopy. We thank N. Wingreen, C. Brangwynne, C. Brody, H. Stone, S. Little, S. Di Talia, Y.-C. Wang and Y. Yan for their suggestions on the manuscript. We thank all members of the Wieschaus and Schupbach laboratories for discussions. This work was supported by the National Institutes of Health (National Institute of Child Health and Human Development grant 5R37HD15587) to E.F.W. and by the Howard Hughes Medical Institute. B.H. was supported by the New Jersey Commission on Cancer Research Fellowship. The Imaging core facility was supported by National Institutes of Health Grant P50 GM 071508.

Author Contributions B.H., K.D., O.P. and E.W. designed the study, performed the experiments and analysed the data. B.H. wrote the first draft of the manuscript. All authors participated in discussion of the data and in producing the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (efw@princeton.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13047

A committed precursor to innate lymphoid cells

Michael G. Constantinides[1], Benjamin D. McDonald[1], Philip A. Verhoef[1] & Albert Bendelac[1]

Innate lymphoid cells (ILCs) specialize in the rapid secretion of polarized sets of cytokines and chemokines to combat infection and promote tissue repair at mucosal barriers[1–9]. Their diversity and similarities with previously characterized natural killer (NK) cells and lymphoid tissue inducers (LTi) have prompted a provisional classification of all innate lymphocytes into groups 1, 2 and 3 solely on the basis of cytokine properties[10], but their developmental pathways and lineage relationships remain elusive. Here we identify and characterize a novel subset of lymphoid precursors in mouse fetal liver and adult bone marrow that transiently express high amounts of PLZF, a transcription factor previously associated with NK T cell development[11,12], by using lineage tracing and transfer studies. PLZF^{high} cells were committed ILC progenitors with multiple ILC1, ILC2 and ILC3 potential at the clonal level. They excluded classical LTi and NK cells, but included a peculiar subset of NK1.1^{+}DX5^{-} 'NK-like' cells residing in the liver. Deletion of PLZF markedly altered the development of several ILC subsets, but not LTi or NK cells. PLZF^{high} precursors also expressed high amounts of ID2 and GATA3, as well as TOX, a known regulator of PLZF-independent NK and LTi lineages[13]. These findings establish novel lineage relationships between ILC, NK and LTi cells, and identify the common precursor to ILCs, termed ILCP. They also reveal the broad, defining role of PLZF in the differentiation of innate lymphocytes.

To study the expression pattern of *Zbtb16* encoding the transcription factor PLZF, which directs the developmental acquisition of the innate effector program of NK T cells[11,12,14,15], we inserted a sequence coding for a fusion of enhanced green fluorescent protein (GFP) and Cre downstream of an IRES after the last exon of *Zbtb16* (Extended Data Fig. 1a). As expected, GFP was selectively expressed in the NK T lineage, with early developmental stages 1 and 2 showing higher levels than mature stage 3 cells, but was not found in the bone marrow common lymphoid precursor (CLP), T cells or B cells of PLZF$^{GFPcre+/-}$ mice (Fig. 1a). In PLZF$^{GFPcre+/-}$ mice carrying the ROSA26-floxstop-yellow fluorescent protein (YFP) fate-mapping allele, nearly all NK T cells expressed YFP, as expected, although approximately 35% of cells in all lymphoid and myeloid lineages were also labelled (Extended Data Fig. 1b, c and data not shown). Because haematopoietic stem cells (HSC) did not express GFP but were already labelled by YFP, this 'background' reflected some expression of PLZF before the HSC stage, probably in multipotent embryonic cells. Indeed, after transfer of fluorescence-activated cell sorting (FACS)-sorted YFP-negative bone marrow cells into lethally irradiated recipients, 94% of NK T cells still expressed YFP, whereas donor-derived CLPs, B cells and T cells were unlabelled (Fig. 1a). Thus, the experiments shown in Fig. 1 were conducted with such bone marrow chimaeras, although all results were confirmed in non-chimaeric reporter mice. Intriguingly, several innate lymphoid lineages, which arise from CLP, were also specifically labelled by YFP despite absence of GFP expression. Thus, ILC2 in the lungs, intestinal lamina propria (LP), peritoneal cavity and mesenteric lymph nodes were all fate-mapped (Fig. 1a, d, Extended Data Fig. 1b and data not shown). Immature ILC2 in the bone marrow (iILC2) expressed very low amounts of GFP, but were already maximally labelled by YFP, suggesting expression of a higher level of PLZF at an earlier stage of their development (Fig. 1a, d). Group 1 innate lymphocyte subsets showed heterogeneous PLZF tracing:

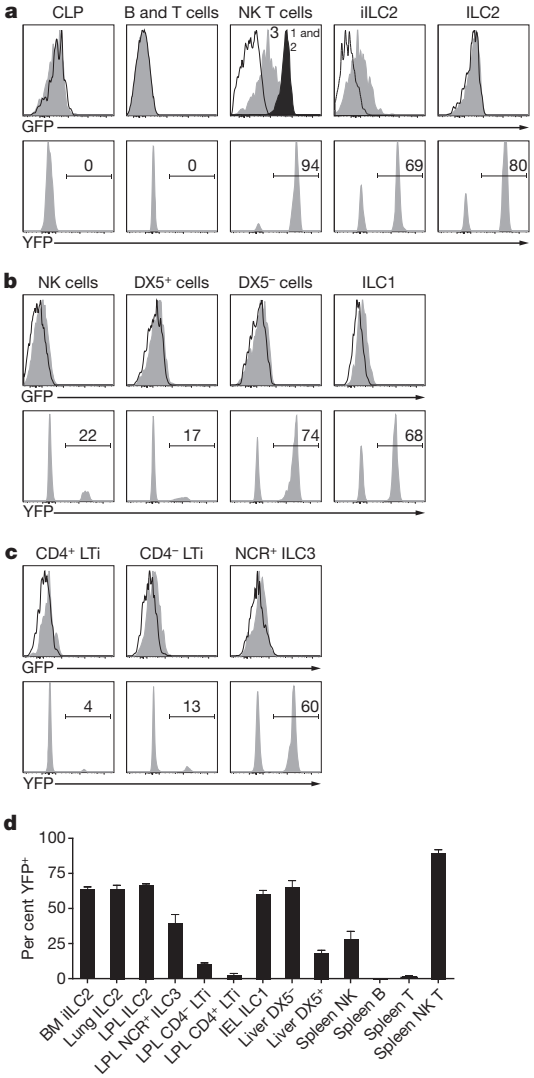


Figure 1 | ILC lineage tracing in PLZF^{GFPcre} reporter mice. a–c, Top rows, expression of GFP by indicated cell-types of PLZF$^{GFPcre+/+}$ mice (filled grey) and wild type (WT; open). Bottom rows, YFP expression in radiation chimaeras reconstituted with YFP^{-}Lin^{-}Sca-1^{+}cKit^{+} (LSK) bone marrow cells from PLZF$^{GFPcre+/-}$ ROSA26-YFP mice. CLP and iILC2 from bone marrow (BM); B and T cells and NK cells from spleen; NK T cells from thymus (GFP), divided into early stages 1, 2 and late stage 3, and from spleen (YFP); ILC2 from lung; DX5^{+} and DX5^{-} NK cells from liver; ILC1 from intestinal intraepithelial lymphocytes (IEL); LTi and ILC3 from LPL. BM iILC2 and lung ILC2 were identified as Lin^{-}CD25^{+}IL-7Rα^{+}Thy1.2^{+}; LPL ILC2 as CD3ε^{-}CD19^{-}CD25^{+}KLRG1^{+}; LPL NCR^{+} ILC3 as CD3ε^{-}CD19^{-}NKp46^{+}NK1.1^{-}; LPL CD4^{-} LTi cells as CD3ε^{-}CD19^{-}IL-7Rα^{+}KLRG1^{-}CCR6^{+}CD4^{-}; LPL CD4^{+} LTi cells as CD3ε^{-}CD19^{-}CCR6^{+}CD4^{+}; and IEL ILC1 as CD3ε^{-}CD19^{-}NKp46^{+}NK1.1^{+}CD160^{+}. FACS identification of the remaining subsets is defined in the methods section. **d**, Summary of results (mean ± s.e.m.). Data representative of 3–9 individual mice analysed in at least 2 independent experiments.

[1]Committee on Immunology, Department of Pathology, The Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637, USA.

©2014 Macmillan Publishers Limited. All rights reserved

whereas few splenic NK cells expressed YFP, intestinal intraepithelial NK-like cells (termed ILC1[1]) were prominently labelled (Fig. 1b). In the liver the recently described non-recirculating DX5$^-$CD49a$^+$ subset of CD3ε$^-$NK1.1$^+$ cells[16] was heavily labelled, whereas classical recirculating DX5$^+$CD49a$^-$ NK cells were mostly negative. Different subsets of group 3 innate lymphocytes in the LP also showed markedly different patterns of tracing (Fig. 1c and Extended Data Fig. 2). CD4$^+$ and CD4$^-$ LTi were not labelled, whereas NCR$^+$ ILC3 prominently expressed YFP. In summary, PLZF lineage-tracing labelled not only ILC2 but also the subsets of group 1 and group 3 cells that are most clearly distinguishable from classical NK and LTi cells, respectively, and will hereafter be termed ILC1 and ILC3.

Searching for the PLZF-expressing precursor of ILCs, we identified a rare subset of PLZF^{high} cells in fetal liver and adult bone marrow. They showed a homogeneous lineage$^-$(Lin$^-$)IL-7Rα$^+$cKit$^+$α4β7^{high} phenotype (Fig. 2a, b), similar to the CLP-derived subset previously suggested to contain precursors for LTi[17–20]. The PLZF^{high} population represented approximately 5% of Lin$^-$IL-7Rα$^+$cKit$^+$ cells and approximately 30% of the α4β7^{high} fraction (Fig. 2b, c). It expressed Thy1 but lacked expression of markers associated with mature ILCs, NK or LTis such as CD4, CXCR6, CD25, IL-17RB, T1/ST2, Sca-1, CD122, NK1.1, CCR6 and NKp46 (Fig. 2c, d and data not shown). Interestingly, the PLZF^{high} cells included a fraction of CD62L^{high}ICOS^{low} cells, probably representing the earliest developmental stage after PLZF expression, because PLZF characteristically induces the downregulation of CD62L and upregulation of ICOS[15,21]. Transcriptional analysis of purified PLZF^{high} cells compared with bone marrow CLP, iILC2 and NK progenitors (NKP), and with ILC3-enriched lamina propria lymphocytes (LPL), confirmed the very high amounts of *Zbtb16* mRNA encoding PLZF, as well as of other key transcription factors such as *Id2*, *Gata3* and *Rora*, which are required for ILC2 development[3,22–25] (Fig. 2e and Extended Data Fig. 3) and *Tcf7*, a target of Notch required for ILC2 and ILC3[26,27]. Notably, they expressed very high amounts of *Tox*, which is required for the development of both NK and LTi cells[13], and low but detectable amounts of *Tbx21* and *Rorc*, the transcription factors associated with ILC1 and ILC3, respectively[1,6]. Single-cell analysis by intracellular flow cytometry was performed after MACS-depletion of Lin$^+$ cells in fetal liver and adult bone marrow, gating on Lin$^-$IL-7Rα$^+$ α4β7^{high} cells (Fig. 2f). PLZF^{high} cells, which represented up to 40% of this population, characteristically co-expressed high amounts of both GATA3 and TOX. Interestingly, a small fraction of these PLZF^{high} cells also expressed RORγt, although these were mostly found in the fetal liver and were rare in adult bone marrow. A distinct population consisting of RORγt^{high} PLZF$^-$ cells, which was more abundant in the fetal liver, co-expressed high levels of TOX but not GATA3, probably representing LTi precursors. Finally, a fraction of Lin$^-$IL-7Rα$^+$α4β7^{high} cells lacked both PLZF and RORγt, but expressed TOX. These may be earlier undifferentiated precursors to the ILC and LTi lineages, and may also include precursors to classical NK cells.

To determine precursor-product relationships, we injected a 1:1 mixture of PLZF^{high} cells and CLP cells into *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice. The PLZF^{high} cells, which were purified from a ROSA-floxstop-tdTomato ×

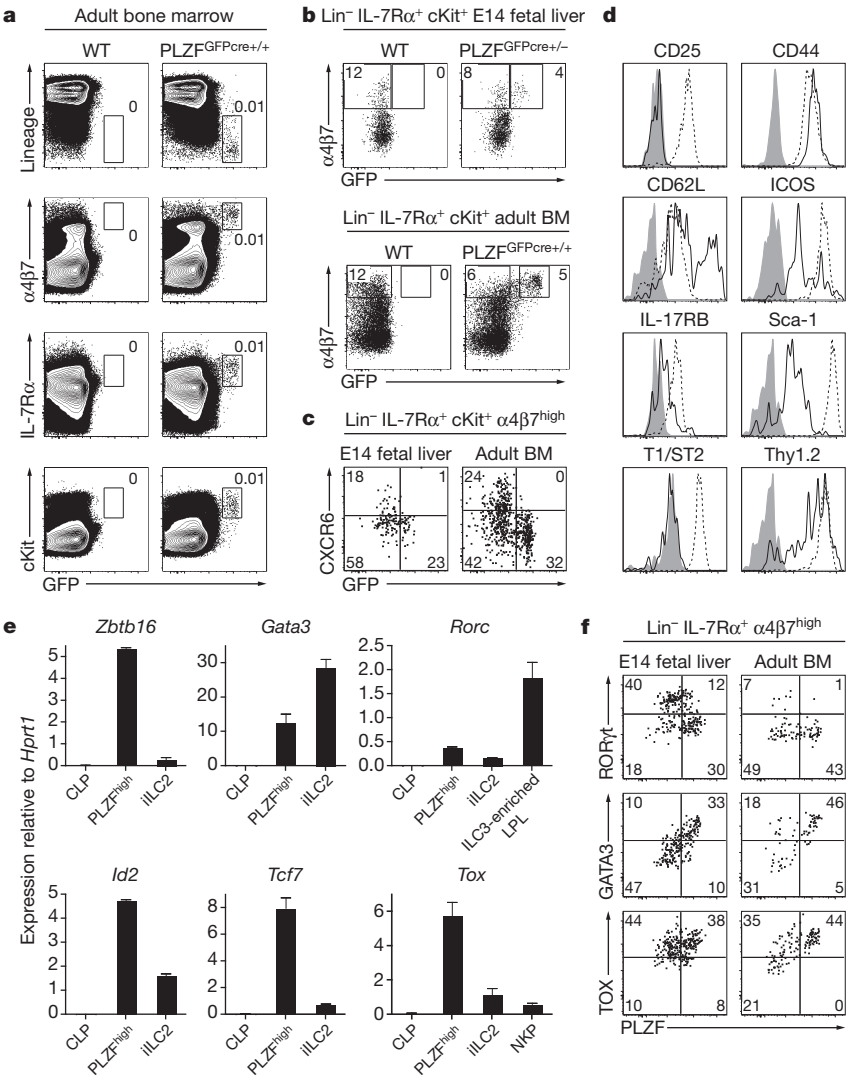
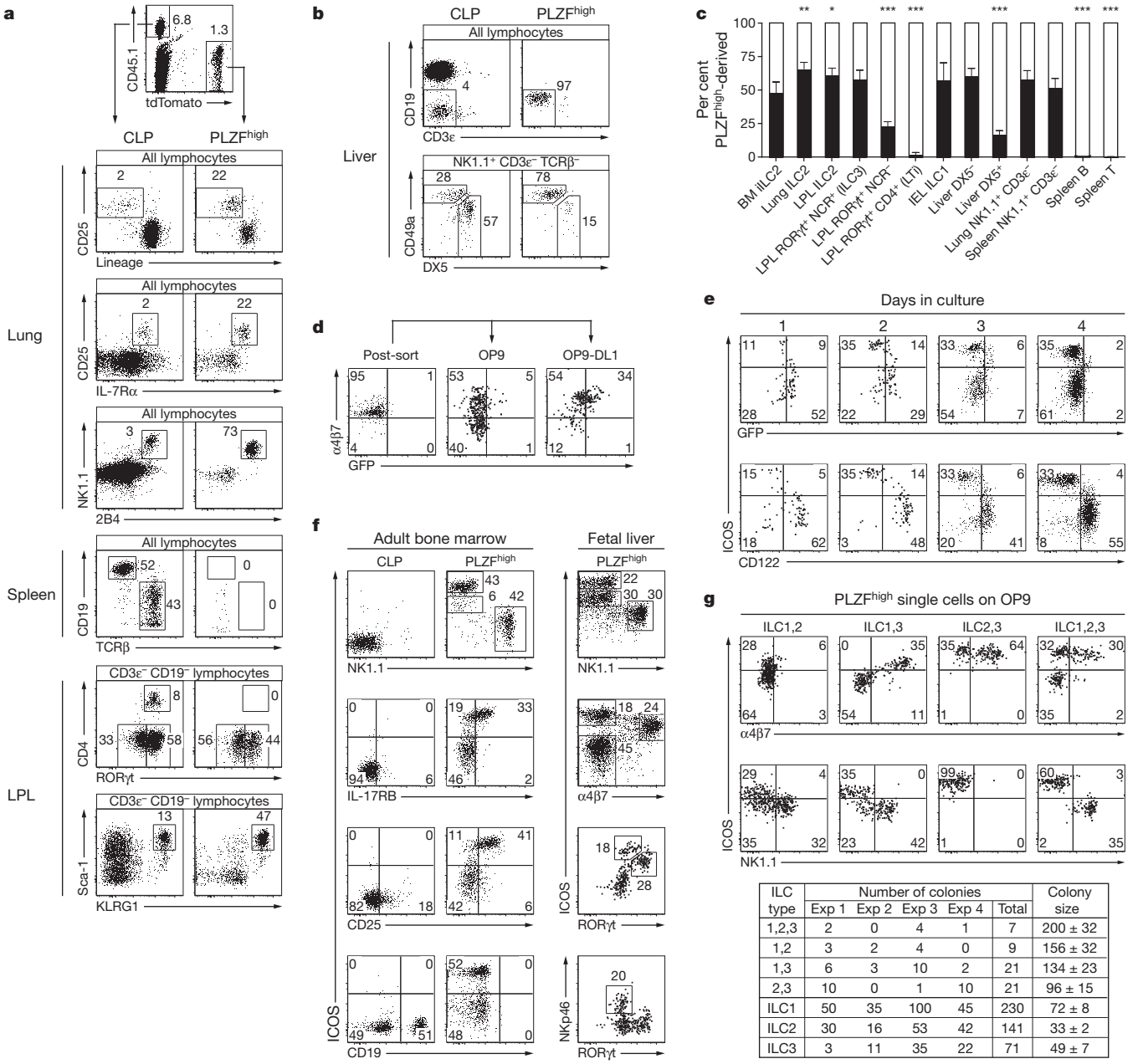


Figure 2 | PLZF^{high} cells in fetal liver and bone marrow. **a**, FACS analysis of total BM cells from adult WT littermate and PLZF$^{GFPcre+/+}$ mice. Data representative of at least 3 independent experiments. **b**, **c**, GFP expression is confined to the α4β7^{high} population of indicated fractions in fetus and adult. Representative of at least 3 independent experiments. **d**, FACS analysis of PLZF^{high} BM cells (solid line) and BM iILC2 (dashed line) compared with isotype controls (filled). iILC2 were gated as Lin$^-$ICOS$^+$T1/ST2$^+$Sca-1$^+$Thy1.2$^+$ for CD25 profile, as CD25$^+$CD62L$^-$Sca-1$^+$Thy1.2$^+$ for CD44 profile, as CD25$^+$CD44$^+$Sca-1$^+$Thy1.2$^+$ for CD62L profile, as CD25$^+$T1/ST2$^+$Sca-1$^+$Thy1.2$^+$ for ICOS profile, as CD25$^+$ICOS$^+$Sca-1$^+$Thy1.2$^+$ for T1/ST2 profile, as CD25$^+$Thy1.2$^+$ for Sca-1 profile or as CD25$^+$ICOS$^+$Sca-1$^+$ for Thy1.2 profile. Representative of at least 3 independent experiments. **e**, Quantitative PCR with reverse transcription analysis in bone marrow subsets as indicated. NKP are Lin$^-$CD27$^+$IL-7Rα$^-$Flt3$^-$CD122$^+$ BM cells and ILC3-enriched LPL are CD3ε$^-$CD19$^-$IL-7Rα$^+$Thy1.2$^+$KLRG1$^-$CD4$^-$ cells. Mean ± s.e.m. of data from 2–3 independent experiments. **f**, Intracellular FACS analysis for transcription factors in cells as indicated. Representative of at least 3 independent experiments.

©2014 Macmillan Publishers Limited. All rights reserved

PLZF$^{GFPcre+/-}$ background to further ascertain their history of PLZF expression based on tdTomato expression, gave rise to a diversity of innate lineages, with the notable exception of CD4$^+$ LTi (Fig. 3a). Significantly more peripheral ILC2 were derived from the PLZF^{high} fraction at 5–7 weeks post-injection (Fig. 3a, c). Both progenitors gave rise to similar amounts of CD3ε$^-$NK1.1$^+$ group 1 cells, but CLP mostly generated the classical DX5$^+$CD49a$^-$NK1.1$^+$ NK cells, whereas the PLZF^{high} cells primarily generated a distinct DX5$^-$CD49a$^+$NK1.1$^+$ ILC1



ILC type	Number of colonies					Colony size
	Exp 1	Exp 2	Exp 3	Exp 4	Total	
1,2,3	2	0	4	1	7	200 ± 32
1,2	3	2	4	0	9	156 ± 32
1,3	6	3	10	2	21	134 ± 23
2,3	10	0	1	10	21	96 ± 15
ILC1	50	35	100	45	230	72 ± 8
ILC2	30	16	53	42	141	33 ± 2
ILC3	3	11	35	22	71	49 ± 7

Figure 3 | PLZF^{high} cells are ILC progenitors. **a**–**c**, CD45.2 *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice were injected with equivalent numbers of tdTomato$^+$ PLZF^{high} cells and CD45.1 CLP (800–1,200 of each) and the progeny of these populations analysed 5–7 weeks later by FACS, as indicated. Summary bar graph of mean percentages ± s.e.m. in **c**, with significant differences between PLZF^{high}- (black bar) and CLP-derived (white bar) shown by *. LPL RORγt$^+$ NCR$^+$ cells were identified as CD3ε$^-$CD19$^-$RORγt$^+$NKp46$^+$CD4$^-$; LPL RORγt$^+$NCR$^-$ cells as CD3ε$^-$CD19$^-$RORγt$^+$NKp46$^-$CD4$^-$; and LPL RORγt$^+$CD4$^+$ cells as CD3ε$^-$CD19$^-$RORγt$^+$CD4$^+$. Remaining populations were gated as indicated in Fig. 1. Data representative of 6–9 chimaeras analysed in at least 2 independent experiments. **d**, FACS analysis of Lin$^-$IL-7Rα$^+$α4β7^{high}GFP$^-$CXCR6$^-$ cells sorted from PLZF$^{GFPcre+/-}$ BM and cultured on OP9 or OP9-DL1 for 48 h. Data representative of 3 independent experiments. **e**, FACS analysis of sorted PLZF^{high} BM cells cultured on OP9 for indicated periods. Data representative of at least 4 replicate cultures for each time point from 2 or more independent experiments. **f**, FACS analysis of PLZF^{high} or CLP cells from adult BM or fetal liver cultured on OP9 for 4 days. Data representative of at least 4 replicate cultures from 2 or more independent experiments. **g**, FACS analysis of representative colonies originating from single fetal liver PLZF^{high} cells that were sorted and cultured into 96-well plates containing irradiated OP9 stromal cells for 5–6 days. ILC1 were characterized as ICOS^{low}α4β7$^-$ NK1.1$^+$ populations, ILC2 as ICOS^{high}α4β7$^-$NK1.1$^-$, and ILC3 as ICOS$^{int/high}$α4β7$^+$NK1.1$^-$. The table summarizes the analysis of thirteen 96-well plates analysed in four independent experiments (three plates in experiments 1, 2 and 4; four plates in experiment 3), with the average colony size ± s.e.m. as indicated. In experiment 4, we mixed a 1:1 ratio of CD45.1/5.2 and CD45.2 PLZF$^{GFPcre+/-}$ fetal liver cells before single-cell sorting. All 122 colonies were either CD45.1/5.2 ($n = 59$) or CD45.2 ($n = 63$), ruling out doublet contamination as an explanation for the presence of mixed ILC colonies ($P < 0.01$ Chi-square). The cloning efficiency was 40% on average, with all but 17 colonies (not shown in the table) unambiguously assigned to defined ILC lineages.

©2014 Macmillan Publishers Limited. All rights reserved

subset in the liver (Fig. 3b, c). PLZF^{high}-derived NK1.1$^+$CD3ε$^-$TCRβ$^-$ cells were observed in the spleen and lungs, and expressed significantly more NKp46 than CLP-derived NK cells (Extended Data Fig. 4). Unlike CLP, PLZF^{high} cells did not generate B cells, consistent with their expression of *Tcf7*, suggesting that they had already received a Notch signal *in vivo*. They also gave rise to far fewer T cells than CLP after transfer into *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice *in vivo* (Fig. 3a, c) or after co-culture with fetal thymic stroma *in vitro* (Extended Data Fig. 5), indicating that PLZF^{high} cells had also largely lost T cell potential. Thus, the PLZF^{high} cell is an ILC precursor (ILCP) heavily committed to innate lymphoid lineages, excluding classical NK and LTi cells.

When cultured on OP9 stromal cells, the GFP$^-$ fraction of Lin$^-$IL-7Rα$^+$cKit$^+$α4β7^{high} cells generated a small fraction of GFP$^+$ cells within 2 days of culture, which was increased upon culture on OP9-DL1 cells (Fig. 3d), confirming the precursor–product relationship and its promotion by Notch signals. To study the short-term fate of ILCP *in vitro*, we sorted Lin$^-$IL-7Rα$^+$cKit$^+$α4β7^{high} PLZF^{high} cells from either adult bone marrow or fetal liver and cultured them on OP9 stromal cells, without Notch ligands, in the presence of non-polarizing IL-7 and stem-cell factor (SCF). By 24 h, a substantial fraction of the PLZF^{high} population had already begun to downregulate GFP and to upregulate either ICOS or CD122 (Fig. 3e), and by 4 days, the cells had resolved into three separate populations of ILC1 (ICOS$^-$IL-17RB$^-$T1/ST2$^-$CD25$^-$RORγt$^-$GATA3$^-$α4β7$^-$CD122$^+$NK1.1$^+$NKp46$^+$), ILC2 (ICOS^{high}IL-17RB$^+$T1/ST2$^+$CD25$^+$RORγt$^-$GATA3$^+$α4β7$^-$CD122$^-$NK1.1$^-$NKp46$^-$), and ILC3 (ICOS^{int}IL17RB$^-$T1/ST2$^-$CD25$^-$RORγt$^+$GATA3$^-$α4β7$^+$CD122$^-$NK1.1$^-$NKp46$^{-/low}$) (Fig. 3f and Extended Data Fig. 6). This transient expression of PLZF before ILC differentiation may explain why fate-mapping by the ROSA-YFP allele in Fig. 1 did not quite reach 100% for these populations as in NK T cells where PLZF expression is permanent. Notably, ILC3 cells consistently arose with higher frequency from fetal liver than from adult bone marrow cells (Fig. 3f). In single-cell cultures with OP9 stromal cells, fetal PLZF^{high} cells generated mixtures of ILC lineages, including triple and double ILC1, ILC2 and ILC3, in 58 out of 500 wells (12%) as well as single ILC1, ILC2 or ILC3 populations in the remaining wells (Fig. 3g). Mixing CD45 allele-marked fetal liver cells before single-cell sorting confirmed that the mixed ILC colonies did not result from cell aggregates escaping exclusion during cell-sorting (Fig. 3g, experiment 4) and post-sorting microscopic assessment confirmed single-cell seeding in each of 96 wells. These results indicated that the PLZF^{high} cells contained a common ILC progenitor characterized by multiple ILC lineage potential at the clonal level, and that subsequent commitment to individual ILC1, ILC2 or ILC3 lineages occurred rapidly during the PLZF^{high} stage of ILC development. Wells containing mixtures of ILC lineages had more growth than those with single lineages, indicating that multipotent ILC progenitors retained greater proliferative capacity.

We then assessed the contribution of PLZF to ILC development in lethally irradiated recipients of a mixture of bone marrow from congenically labelled *Zbtb16*$^{-/-}$ and *Zbtb16*$^{+/+}$ littermate donors (Fig. 4a). The production of ILC2 was markedly decreased in the PLZF-deficient compartment of all tissues, demonstrating a cell-intrinsic requirement of PLZF. Furthermore, residual immature ILC2 in the bone marrow expressed significantly less ICOS and more CD62L in the absence of PLZF (Fig. 4b, c), a phenotype entirely consistent with that of PLZF-deficient NK T cells[11,12,15], which accounts for altered recirculating properties and reduced frequencies in peripheral tissues. LTi cells were unaffected by the genetic deletion of PLZF, as expected (Fig. 4a). Other ILC3 cells, including NCR$^+$ ILC3, were unaffected, despite their expression of PLZF during development. However, liver ILC1 (DX5$^-$CD49a$^+$) were profoundly affected, similar to ILC2, whereas classical (DX5$^+$CD49a$^-$) NK cells were not (Fig. 4a).

Our studies have identified the elusive CLP-derived common progenitor to the ILC1, ILC2, and ILC3 lineages and demonstrated close but distinct lineage relationships with classical NK and LTi cells. We have shown that a committed PLZF^{high} ILCP emerges from within the same IL-7Rα$^+$α4β7^{high} population that includes precursors to LTi and possibly also to NK cells (Extended Data Fig. 7). It is uniquely distinguished and affected by the rapid and transient upregulation of PLZF, whereas TOX is more broadly induced and reciprocally controls classical NK and LTi cells. The concomitant induction of GATA3, RORα, Tbet and RORγt, which govern the polarization of ILCs, parallels the sequence during PLZF-dependent NK T cell differentiation[28], indicating a broad and defining role of PLZF in the differentiation of innate lymphocyte lineages.

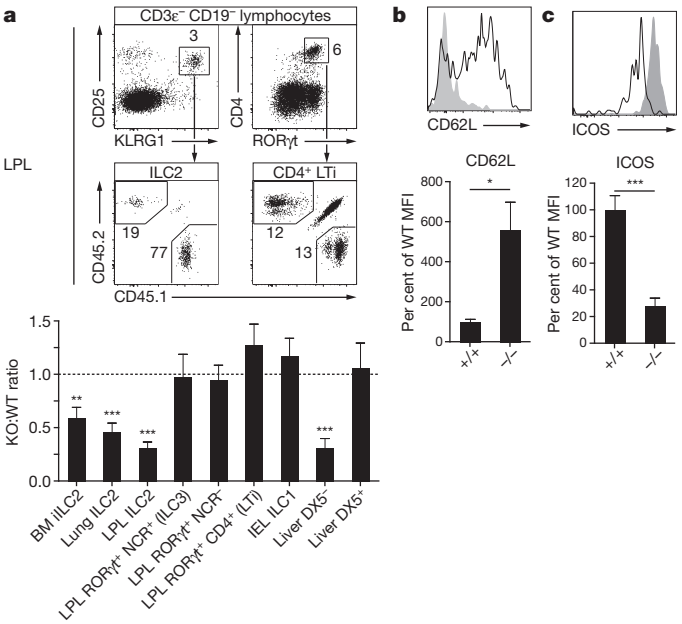


Figure 4 | PLZF is required for ILC development. Lethally irradiated CD45.1/2 mice were reconstituted with a mixture of CD45.2 *Zbtb16*$^{-/-}$ and CD45.1 WT BM and analysed as indicated 5–7 weeks later. **a**, Representative FACS analysis of LPL gated as indicated, and summary bar graph of mean percentages ± s.e.m. for indicated populations, with significant variation from 1.00 indicated by *. Populations were gated as indicated in Fig. 3. Data representative of 6–10 chimaeras analysed in at least 2 independent experiments. **b**, **c**, FACS analysis of iILC2 from *Zbtb16*$^{-/-}$ mice and *Zbtb16*$^{+/+}$ littermates, representative of 5 (**b**) and 8 (**c**) mice of each genotype analysed in 3 or more independent experiments.

METHODS SUMMARY

PLZF^{GFPcre} mice were generated by inserting a sequence encoding an IRES and EGFP–Cre fusion protein immediately after the *Zbtb16* stop codon through homologous recombination in C57BL/6J embryonic stem cells. For fate-mapping, they were crossed to mice carrying a ROSA26-fl-STOP-fl-YFP or a ROSA26-fl-STOP-fl-tdTomato allele. Bone marrow chimaeras were generated by reconstituting lethally irradiated recipients with purified Lin$^-$ Sca-1$^+$ cKit$^+$ (LSK) bone marrow cells. PLZF^{high} ILC precursors were purified from bone marrow or fetal liver and cultured in bulk or as single cells deposited by flow cytometry in 96-well plates with irradiated OP9 or OP9-DL1 stromal cells and murine IL-7 and SCF. For *in vivo* transfers, PLZF^{high} ILC precursors and CLP were purified from bone marrow and injected into *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice. Quantitative PCR with reverse transcription for transcription factors was performed using TaqMan primers.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2013; accepted 21 January 2014.
Published online 9 February 2014.

1. Fuchs, A. *et al.* Intraepithelial type 1 innate lymphoid cells are a unique subset of IL-12- and IL-15-responsive IFN-γ-producing cells. *Immunity* **38,** 769–781 (2013).
2. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464,** 1367–1370 (2010).
3. Moro, K. *et al.* Innate production of T_H2 cytokines by adipose tissue-associated c-Kit$^+$Sca-1$^+$ lymphoid cells. *Nature* **463,** 540–544 (2010).

©2014 Macmillan Publishers Limited. All rights reserved

4. Price, A. E. *et al.* Systemically dispersed innate IL-13-expressing cells in type 2 immunity. *Proc. Natl Acad. Sci. USA* **107,** 11489–11494 (2010).
5. Monticelli, L. A. *et al.* Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nature Immunol.* **12,** 1045–1054 (2011).
6. Satoh-Takayama, N. *et al.* Microbial flora drives interleukin 22 production in intestinal NKp46$^+$ cells that provide innate mucosal immune defense. *Immunity* **29,** 958–970 (2008).
7. Sanos, S. L. *et al.* RORgammat and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46$^+$ cells. *Nature Immunol.* **10,** 83–91 (2009).
8. Luci, C. *et al.* Influence of the transcription factor RORγt on the development of NKp46$^+$ cell populations in gut and skin. *Nature Immunol.* **10,** 75–82 (2009).
9. Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457,** 722–725 (2009).
10. Spits, H. *et al.* Innate lymphoid cells–a proposal for uniform nomenclature. *Nature Rev. Immunol.* **13,** 145–149 (2013).
11. Savage, A. K. *et al.* The transcription factor PLZF directs the effector program of the NKT cell lineage. *Immunity* **29,** 391–403 (2008).
12. Kovalovsky, D. *et al.* The BTB-zinc finger transcriptional regulator PLZF controls the development of invariant natural killer T cell effector functions. *Nature Immunol.* **9,** 1055–1064 (2008).
13. Aliahmad, P., de la Torre, B. & Kaye, J. Shared dependence on the DNA-binding factor TOX for the development of lymphoid tissue-inducer cell and NK cell lineages. *Nature Immunol.* **11,** 945–952 (2010).
14. Savage, A. K., Constantinides, M. G. & Bendelac, A. Promyelocytic leukemia zinc finger turns on the effector T cell program without requirement for agonist TCR signaling. *J. Immunol.* **186,** 5801–5806 (2011).
15. Constantinides, M. G., Picard, D., Savage, A. K. & Bendelac, A. A naive-like population of human CD1d-restricted T cells expressing intermediate levels of promyelocytic leukemia zinc finger. *J. Immunol.* **187,** 309–315 (2011).
16. Peng, H. *et al.* Liver-resident NK cells confer adaptive immunity in skin-contact inflammation. *J. Clin. Invest.* **123,** 1444–1456 (2013).
17. Yoshida, H. *et al.* Expression of $\alpha_4\beta_7$ integrin defines a distinct pathway of lymphoid progenitors committed to T cells, fetal intestinal lymphotoxin producer, NK, and dendritic cells. *J. Immunol.* **167,** 2511–2521 (2001).
18. Sawa, S. *et al.* Lineage relationship analysis of RORγt$^+$ innate lymphoid cells. *Science* **330,** 665–669 (2010).
19. Possot, C. *et al.* Notch signaling is necessary for adult, but not fetal, development of RORγt$^+$ innate lymphoid cells. *Nature Immunol.* **12,** 949–958 (2011).
20. Cherrier, M., Sawa, S. & Eberl, G. Notch, Id2, and RORγt sequentially orchestrate the fetal development of lymphoid tissue inducer cells. *J. Exp. Med.* **209,** 729–740 (2012).
21. Gleimer, M., von Boehmer, H. & Kreslavsky, T. PLZF controls the expression of a limited number of genes essential for NKT cell function. *Front. Immunol.* **3,** 374 (2012).
22. Hoyler, T. *et al.* The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. *Immunity* **37,** 634–648 (2012).
23. Mjösberg, J. *et al.* The transcription factor GATA3 is essential for the function of human type 2 innate lymphoid cells. *Immunity* **37,** 649–659 (2012).
24. Wong, S. H. *et al.* Transcription factor RORα is critical for nuocyte development. *Nature Immunol.* **13,** 229–236 (2012).
25. Halim, T. Y. *et al.* Retinoic-acid-receptor-related orphan nuclear receptor alpha is required for natural helper cell development and allergic inflammation. *Immunity* **37,** 463–474 (2012).
26. Yang, Q. *et al.* T cell factor 1 is required for group 2 innate lymphoid cell generation. *Immunity* **38,** 694–704 (2013).
27. Mielke, L. A. *et al.* TCF-1 controls ILC2 and NKp46$^+$RORγt$^+$ innate lymphocyte differentiation and protection in intestinal inflammation. *J. Immunol.* **191,** 4383–4391 (2013).
28. Constantinides, M. G. & Bendelac, A. Transcriptional regulation of the NKT cell lineage. *Curr. Opin. Immunol.* **25,** 161–167 (2013).

Acknowledgements We thank W. Yokoyama for discussion; F. Gounari and R. de Pooter for advice on OP9 cultures; H. Gudjonson for statistical advice; D. Leclerc, J. Cao, M. Olsen and R. Duggan for help with cell sorting; V. Bindokas and R. Mathew for help with fluorescence microscopy. This work was supported by NIH grants R01HL118092, R01AI038339 and P30DK42086 and by The Howard Hughes Medical Institute (A.B.).

Author Contributions M.G.C., B.D.M. and P.A.V. designed research, performed experiments and analysed data. M.G.C. and A.B. wrote the paper. A.B. supervised the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. (abendela@bsd.uchicago.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13027

Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430

Travis K. Warren[1], Jay Wells[1], Rekha G. Panchal[1], Kelly S. Stuthman[1], Nicole L. Garza[1], Sean A. Van Tongeren[1], Lian Dong[1], Cary J. Retterer[1], Brett P. Eaton[1], Gianluca Pegoraro[1], Shelley Honnold[1], Shanta Bantia[2], Pravin Kotian[2], Xilin Chen[2], Brian R. Taubenheim[2]†, Lisa S. Welch[1], Dena M. Minning[3], Yarlagadda S. Babu[2], William P. Sheridan[2] & Sina Bavari[1]

Filoviruses are emerging pathogens and causative agents of viral haemorrhagic fever. Case fatality rates of filovirus disease outbreaks are among the highest reported for any human pathogen, exceeding 90% (ref. 1). Licensed therapeutic or vaccine products are not available to treat filovirus diseases. Candidate therapeutics previously shown to be efficacious in non-human primate disease models are based on virus-specific designs and have limited broad-spectrum antiviral potential. Here we show that BCX4430, a novel synthetic adenosine analogue, inhibits infection of distinct filoviruses in human cells. Biochemical, reporter-based and primer-extension assays indicate that BCX4430 inhibits viral RNA polymerase function, acting as a non-obligate RNA chain terminator. Post-exposure intramuscular administration of BCX4430 protects against Ebola virus and Marburg virus disease in rodent models. Most importantly, BCX4430 completely protects cynomolgus macaques from Marburg virus infection when administered as late as 48 hours after infection. In addition, BCX4430 exhibits broad-spectrum antiviral activity against numerous viruses, including bunyaviruses, arenaviruses, paramyxoviruses, coronaviruses and flaviviruses. This is the first report, to our knowledge, of non-human primate protection from filovirus disease by a synthetic drug-like small molecule. We provide additional pharmacological characterizations supporting the potential development of BCX4430 as a countermeasure against human filovirus diseases and other viral diseases representing major public health threats.

Members of the family *Filoviridae* include Ebola virus (EBOV), Marburg virus (MARV), Ravn virus (RAVV), Sudan virus (SUDV) and Bundibugyo virus (BDBV), all of which cause severe viral haemorrhagic fevers in humans. In nature, filoviruses are transmitted by physical contact between infected individuals, presumably via infected bodily fluids[2]. Initial filovirus disease manifestations include fever, headache, vomiting and diarrhoea[3]. Fatal cases are characterized by viraemia, elevated liver-associated enzyme levels, coagulopathy and haemorrhage. Filovirus disease outbreaks occur sporadically, most frequently in sub-Saharan Africa, with reported case fatality rates exceeding 90% (ref. 1). In 2012, simultaneous outbreaks involved MARV, SUDV and BDBV, an emergent filovirus isolated in 2007[4]. The historical occurrence of independent and simultaneous emergence of distinct filoviruses highlights the need for the identification and development of an efficacious broad-spectrum antiviral product.

Although no licensed antiviral preventative or therapeutic agents are currently available to combat filovirus disease in humans, a number of candidates are being developed and have been evaluated in non-human primate filovirus disease models. These models closely reproduce the known clinical and pathophysiological aspects of fatal human infection. Nucleic-acid-based products, antibody therapies and therapeutic vaccines have successfully protected non-human primates from filovirus disease[5–10], but these approaches rely on virus-specific designs that inherently limit the spectrum of activity and potential utility of individual treatments. The development of a single therapeutic agent active against multiple filoviruses would provide a key, cost-effective component of public-health preparedness plans in outbreak-prone regions.

The broad-spectrum antiviral agent ribavirin, a trizole nucleoside effective against multiple pathogenic RNA viruses, is not active against filoviruses[11]. Other small molecules—including the adenosine analogue 3-deazaneplanocin A (c^3-Npc A) and T-705 (favipiravir), a substituted pyrazine compound —have conferred a high-degree of protection against filoviruses in rodents but have not been reported to protect non-human primates[12–18].

BCX4430, a novel nucleoside analogue (Fig. 1a), was synthesized (Supplementary Information) as part of a small-molecule library designed as inhibitors of viral RNA polymerase activity. BCX4430 is designed to inhibit viral RNA polymerase activity indirectly through non-obligate RNA chain termination, a mechanism requiring anabolism of the parent compound to BCX4430-triphosphate (BCX4430-TP). Then, after pyrophosphate cleavage, incorporation of BCX4430-monophosphate (BCX4430-MP) into nascent viral RNA strands would be expected to cause premature termination of transcription and replication of viral RNA. In support of this proposed mechanism, BCX4430 is rapidly phosphorylated to BCX4430-TP in cultured cell lines and primary hepatocytes, similar to the natural adenosine nucleoside (Fig. 1b and Extended Data Fig. 1a). Addition of BCX4430 reduces expression of green fluorescent protein (GFP) in an artificial EBOV minigenome replicon assay (Fig. 1c and Extended Data Fig. 1b), in which virion structural proteins comprising the viral ribonucleoprotein complex mediate transcription and replication of an RNA replicon template containing a GFP-reporter cassette. BCX4430-TP inhibits hepatitis C virus (HCV) RNA polymerase transcriptional activity in a cell-free, isolated enzyme transcription assay (Fig. 1d) (an isolated filovirus RNA polymerase enzyme assay has yet to be reported) and induces premature termination of RNA chain synthesis by HCV RNA polymerase during template-directed primer-extension assays (Fig. 1e and Extended Data Fig. 1c). In virus-infected cells, BCX4430 reduces surface-expressed MARV and EBOV glycoprotein and reduces the production of intracellular and extracellular MARV RNA (Fig. 1f–h and Extended Data Fig. 1d). Additionally, HeLa cells incubated with $\geq$25 μM BCX4430 produce no detectable infectious MARV virus (concentration providing 90% inhibition (IC_{90}) = 5.4 μM) (Fig. 1h).

Taken together, these assessments strongly support our hypothesis that BCX4430 inhibits viral RNA polymerase function by inducing RNA chain termination. Findings from primer-extension reactions suggest that termination occurs two bases after incorporation of BCX4430-MP, perhaps as a result of inhibitory stereochemical distortions of the nascent RNA chain. We observed no evidence of BCX4430-MP incorporation into human RNA or DNA on exposing human Huh-7 cells to concentrations of BCX4430 exceeding the MARV IC_{50} by more than tenfold (Extended Data Fig. 1f). The basis of the selectivity of BCX4430 for viral polymerases is not yet known.

[1]Division of Molecular and Translational Sciences, Therapeutic Discovery Center, United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Fort Detrick, Maryland 21702, USA. [2]BioCryst Pharmaceuticals Inc., Durham, North Carolina 27703, USA. [3]MedExpert Consulting, Inc., Indialantic, Florida 32903, USA. †Present address: Wilco Consulting, LLC, Durham, North Carolina 27712, USA.

©2014 Macmillan Publishers Limited. All rights reserved

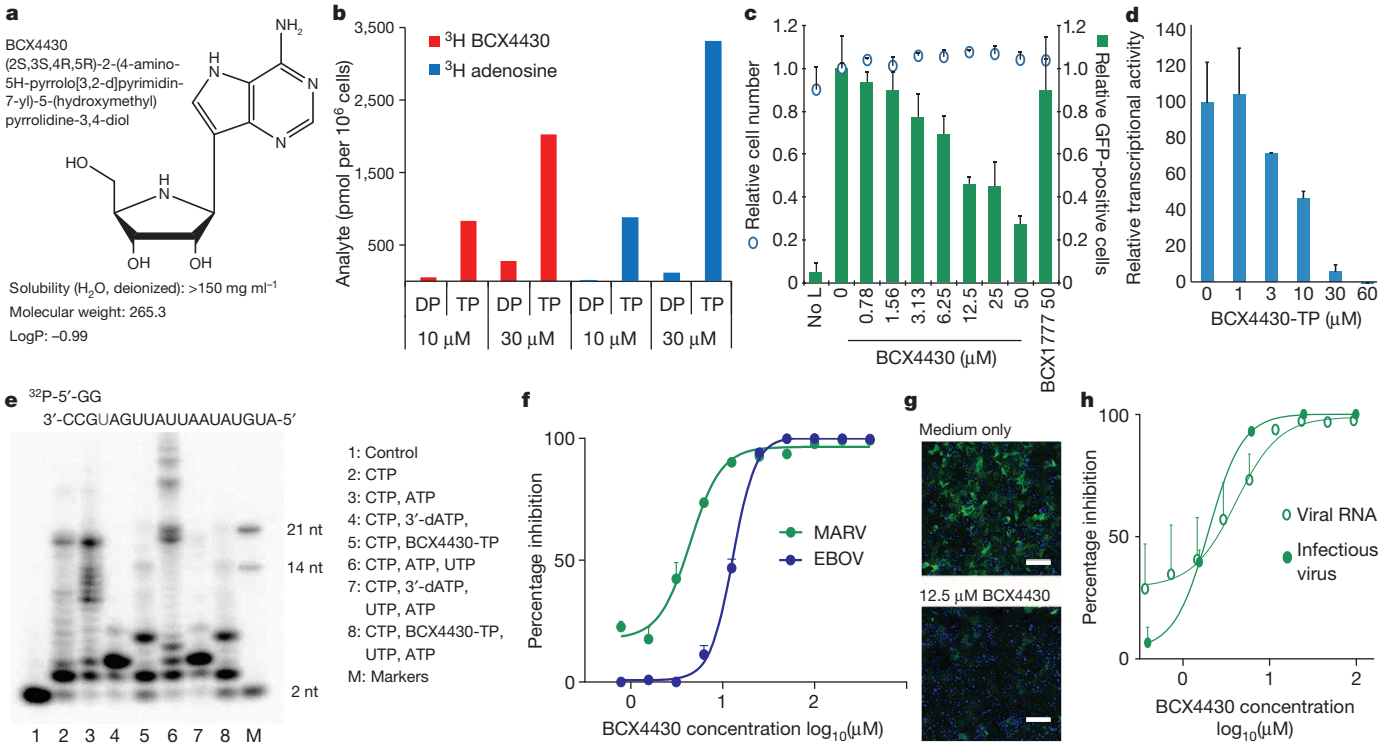


Figure 1 | Pharmacological characterization of BCX4430. **a**, BCX4430 chemical structure and properties. LogP, log of partition coefficient. **b**, Conversion of ^{3}H-BCX4430 or ^{3}H-adenosine to diphosphate (DP) or triphosphate (TP) forms in Huh-7 cells ($n = 1$). **c**, Effect of BCX4430 on replication of an EBOV minigenome RNA replicon in cultured cells ($n = 6$). No L, L-polymerase encoding plasmid omitted from reaction. **d**, BCX4430-TP suppresses HCV NS5B (NS5B-1b-Δ21) RNA polymerase activity ($n = 2$). **e**, RNA products synthesized by purified HCV polymerase, in a template-directed primer (^{32}P-5′-GG) extension assay. nt, nucleotides. **f**, Expression of EBOV and MARV glycoprotein in infected HeLa cells ($n = 5$–6). **g**, MARV-infected HeLa cells. Green, α-MARV glycoprotein; blue, Hoechst dye. Scale bar, 100 μm. **h**, Production of intracellular MARV RNA ($n = 4$) and infectious virus in MARV-infected HeLa cells ($n = 2$ technical replicates). Data in **c**, **d**, and **h** are expressed as mean + standard deviation (s.d.). Data in **f** are expressed as mean + standard error of the mean (s.e.m.).

To assess the spectrum of antiviral activity of BCX4430, we conducted additional high-content image (HCI)-based and neutral-red uptake antiviral assays using an array of other human viruses. BCX4430 exhibited antiviral activity against negative-sense RNA viruses representing the *Filoviridae*, *Arenaviridae*, *Bunyaviridae*, *Orthomyxoviridae*, *Picornaviridae* and *Paramyxoviridae* families, and positive-sense RNA viruses of the *Flaviviridae* and *Coronaviridae* families (Table 1 and Extended Data Fig. 2). Antiviral effects were specific and did not result from cellular toxicity or anti-proliferative effects, as assessed by enumeration of live cells during all HCI assays and by the lactate dehydrogenase (LDH) release assay. In HeLa cells, the 50% cytotoxic concentration (CC_{50}) of BCX4430 exceeded 500 μM as determined by the LDH assay, and consistently exceeded 100 μM using the HCI-based assay (Table 1). An antiviral effect was confirmed in infected human macrophages cultured from normal peripheral blood monocytes, using EBOV (Extended Data Fig. 1e). It is important to note that immortalized cell lines exhibit a substantially attenuated capacity to convert BCX4430 to its active TP form, compared with primary hepatocytes (Extended Data Fig. 1a). Thus, the results obtained from antiviral assessments relying on immortalized cells may underrepresent the antiviral potency of BCX4430 achievable *in vivo* during experimental or clinical infections.

In preparation for *in vivo* pharmacokinetics and efficacy testing, we sought to further characterize potential adverse pharmacological features of BCX4430 that could limit its utility *in vivo* or clinically. BCX4430 exhibited no mutagenicity when tested using the Ames assay at 5 mg per plate, produced no detectable chromosomal aberrations in human lymphocytes at a concentration of 787 μM, and did not inhibit the hERG ion channel current at 30 μM (data not shown). The compound is metabolically stable (half-life ($t_{1/2}$) > 54 min) when incubated with S9 liver fractions obtained from various animal species (Extended Data Table 1). In mouse, rat, guinea pig and cynomolgus macaque, BCX4430 pharmacokinetics were characterized by rapid clearance from the plasma with a half-life of <5 min (Fig. 2b, Extended Data Fig. 4c and Extended

Table 1 | BCX4430 antiviral activity

Virus family	Virus	Strain/variant	EC_{50} (μM)	EC_{90} (μM)	CC_{50} (μM)
Filoviridae	MARV	Musoke*	4.4	10.5	242
	MARV	Ci67*	6.7	16.1	255
	MARV	Angola*	5.0	12.3	242
	EBOV	Kikwit*	11.8	25.4	>100
	SUDV	Boniface*	3.4	10.3	>100
Togaviridae	VEEV	SH3*	>100	>100	>100
	EEEV	FL93-939†	43.2	>100	>100
	WEEV	California†	21.3	>30	>100
	CHIKV	AF 15561*	>100	>100	>100
Bunyaviridae	RVFV	ZH5018*	41.6	98.0	>100
	LACV	Wisconsin 1960†	13.4	65.0	>100
	MPRLV	HV97021050†	40.1	95.0	>250
Arenaviridae	LASV	Josiah*	43.0	>100	>100
	JUNV	Romero*	42.2	>100	>100
Paramyxoviridae	NiV	Malaysia*	41.9	>100	>100
	RSV	2305†	11.0	25.7	>89
	MeV	Chicago†	6.19	34.4	>296
Coronaviridae	MERS-CoV	Jordan N3*	68.4	>100	>100
	SARS-CoV	Urbani†	57.7	>95	>296
Orthomyxoviridae	Influenza	pH1N1†	10.7	17.0	>296
Picornaviridae	HRV2	HGP†	3.4	45.2	>296
Flaviviridae	YFV	17D†	14.1	46.8	>100
	JEV	SA14†	43.6	93.4	>100
	DENV2	New Guinea C†	32.8	89.3	>296

EC_{50}, 50% effective concentration; EC_{90}, 90% effective concentration. Definitions for virus abbreviations are provided in Methods.
* Antiviral activity assessed by high-content image analysis.
† Antiviral activity assessed by neutral-red uptake assay.

©2014 Macmillan Publishers Limited. All rights reserved

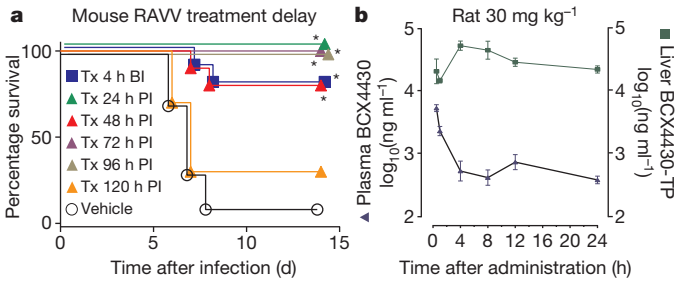


Figure 2 | BCX4430 *in vivo* efficacy and pharmacokinetics characterization. **a**, Survival of RAVV-infected mice ($n = 9$–10 per group). BCX4430 treatments (Tx) were administered i.m. twice daily at 150 mg kg^{-1} doses for 9 days beginning at the indicated time before infection (BI) or post-infection (PI). **b**, Pharmacokinetics of BCX4430 and BCX4430-TP in Sprague–Dawley rats ($n = 3$) after single-dose i.m. administration of BCX4430. *$P < 0.05$ treatment versus vehicle survival curves by log-rank (Mantel–Cox) test. Data in **b** are expressed as mean ± s.e.m.

Data Table 2). By contrast, the half-life of the active BCX4430-TP form in the liver in rats was substantially longer at 6.2 h (Fig. 2b).

Rodent models of filovirus disease do not always reliably reproduce the pathophysiology observed in humans and non-human primates. Nevertheless, rodent systems can provide proof-of-concept efficacy and inform dose and treatment-regimen selection for non-human primate efficacy trials. In RAVV-infected mice, intramuscular (i.m.) administration of BCX4430 conferred a high degree of protection (75–100%) in groups treated prophylactically at doses ≥3.3 mg kg^{-1} (Extended Data Fig. 3a, b). Notably, significant protection ($P < 0.05$) was also observed with doses of 150 mg kg^{-1} initiated at times as long as 96 h after infection (Fig. 2a). Additionally, we demonstrated that mice administered BCX4430 either i.m. or orally were protected against lethal EBOV challenge and that mice receiving i.m. treatments were protected against lethal Rift Valley fever virus challenge (Extended Data Fig. 3c, d).

The *in vivo* post-exposure efficacy of BCX4430 was further verified in two guinea pig models of MARV disease[19] in which animals were challenged either by intraperitoneal (i.p.) injection, or, because filoviruses can be transmitted via aerosolized particles[20], by exposure to aerosolized virus. In both the i.p. and aerosolized-virus challenge models, BCX4430 conferred significant post-exposure protection when treatment was initiated within 48 h of virus exposure (Extended Data Fig. 4a). In the respiratory exposure model, BCX4430 protected guinea pigs when treatment initiation was delayed to 72 h after infection, the longest delay tested (Extended Data Fig. 4b).

We further explored the post-exposure efficacy of BCX4430 using the cynomolgus macaque MARV disease model, which accurately reproduces filovirus disease manifestations observed in fatal human cases[21]. Cynomolgus macaques were experimentally infected with a lethal dose of wild-type MARV (Musoke variant), derived from a human clinical isolate[22], and animals were administered with 15 mg kg^{-1} BCX4430 twice daily by i.m. injection beginning 1–48 h after infection and continuing for 14 days. The six infection-control subjects succumbed by day 12 (Fig. 3a), after having developed viraemia (Fig. 3b) and characteristic signs of filovirus disease, including behavioural inactivity, maculopapular rash, increases in liver injury markers such as aspartate aminotransferase and bilirubin (Fig. 3c, d), and prolongations of prothrombin time (PT) and activated partial thromboplastin time (aPTT) (Fig. 3e, f). All animals treated with BCX4430 beginning 24 or 48 h after infection survived. Five out of six (83%) animals treated with BCX4430 beginning 1 h after infection survived (Fig. 3a). Consistent with the proposed antiviral mechanism of action, BCX4430 significantly and substantially reduced serum MARV burden (Fig. 3b), without inducing type I interferon responses (Extended Data Fig. 4d), as has been observed for c^3-Npc A in mice[15]. Additionally, BCX4430 ameliorated haemorrhagic disease manifestations, as evidenced by shorter PT and aPTT times and reduced laboratory

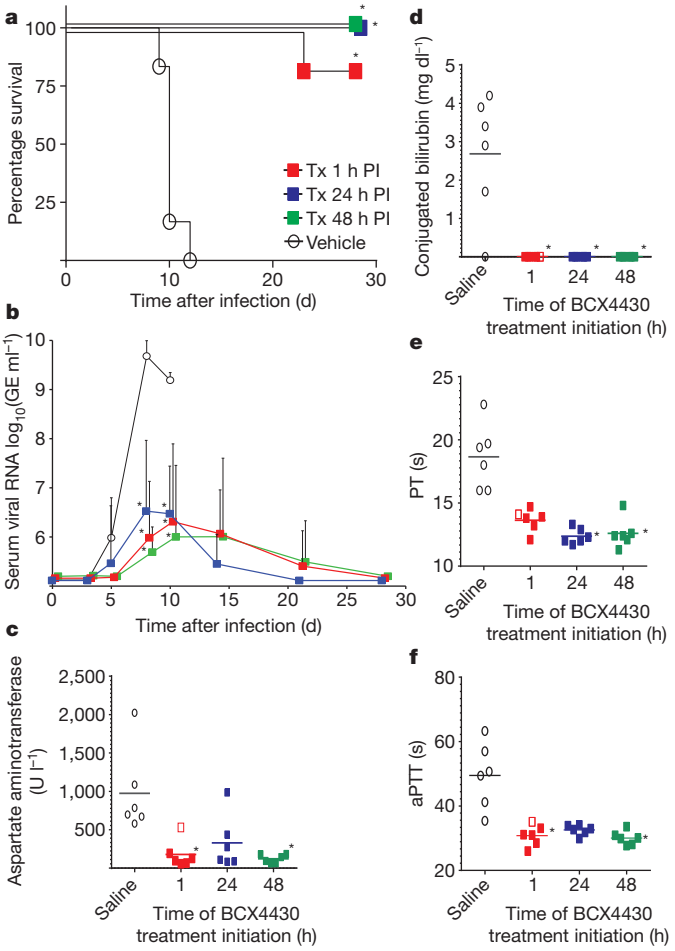


Figure 3 | Post-exposure protection of MARV-infected cynomolgus macaques by BCX4430. **a**–**f**, Animals ($n = 6$) were challenged with MARV by subcutaneous injection, and BCX4430 (Tx) (15 mg kg^{-1} twice daily) or vehicle was administered i.m. beginning at the indicated times after challenge. **a**, Kaplan–Meier survival curves. PI, post-infection. **b**, Serum viral RNA load. GE, genomic equivalents. **c**–**f**, Individual animal maximal values of serum aspartate aminotransferase concentrations (**c**), conjugated bilirubin concentrations (**d**), PT (**e**) and aPTT (**f**). Non-survivors are represented by open symbols. *$P < 0.05$ for comparison of treatment versus vehicle by log-rank (Mantel–Cox) test (**a**), two-tailed analyses using the Holm–Sidak method (**b**), or two-tailed Kruskal–Wallis test followed by Dunn's post-test comparison (**c**–**f**). Data in **b** are expressed as geometric mean + s.d. Horizontal bars in **c**–**f** represent group means.

indices of liver damage, such as serum AST and bilirubin concentrations, compared with vehicle treatment (Fig. 3c–f).

Outbreaks in 2012 involving MARV, SUDV and BDBV reported in Uganda and the Democratic Republic of the Congo highlight the urgent need for development of an effective antiviral product to counter filovirus disease. For the first time, to our knowledge, we report the identification of a small molecule with efficacy against filovirus disease in non-human primates. Furthermore, we have provided evidence that BCX4430 exhibits broad-spectrum antiviral activity against other highly virulent RNA viruses (Table 1 and Extended Data Fig. 2). Additional evaluations are in progress to assess the *in vivo* efficacy of BCX4430 against EBOV and other highly virulent pathogens in non-human primates or other disease models that most closely recapitulate human disease.

BCX4430 was well tolerated, producing no overt signs of systemic toxicity or adverse local reactions in any of the efficacy studies. The substantial efficacy of BCX4430 observed with the i.m. route, which provides high bioavailability and rapid absorption, is conducive for use during outbreaks, potentially enabling administration by individuals lacking advanced medical training. In MARV-infected cynomolgus macaques,

©2014 Macmillan Publishers Limited. All rights reserved

we did not observe any diminution of protective effects with delay of BCX4430 treatment initiation up to 48 h after infection, suggesting that even greater delays of treatment, yet to be tested, may also yield significant protective benefit. In preparation for advancing BCX4430 into human phase 1 clinical trials in the United States, additional studies are in progress to support filing of an Investigational New Drug Application.

METHODS SUMMARY

BCX4430 was synthesized by BioCryst Pharmaceuticals. Cell-based infection assays were conducted using HCI-based analyses as described previously[23] for filoviruses and Rift Valley fever virus. For other viruses, inhibition of virus-induced cytopathic effect (CPE) was assessed in cell-based assays using neutral-red uptake, as described previously[24]. Replication of EBOV minigenome RNA constructs was assayed by using a plasmid-based reconstituted replication/transcription system[25]. Inhibition of RNA transcriptional activity was assessed in an isolated HCV polymerase assay[26] and inhibition of HCV polymerase RNA synthesis activity was evaluated using a template-driven ^{32}P-GG primer extension assay. Mouse models of EBOV, MARV and RAVV disease have been described previously[27,28]. Experiments were conducted using mouse-adapted strains of EBOV (Mayinga variant) and RAVV (Ravn variant). A guinea pig MARV-Musoke parenteral challenge model and virus aerosolization methods used for guinea pig exposures have been previously described[19,29]. BCX4430 was administered in a vehicle of sterile 0.9% saline for injection for all *in vivo* applications. Animal infection experiments were performed in biosafety level 4 containment facilities at the United States Army Medical Research Institute of Infectious Diseases (USAMRIID). All experimental treatment replication was conducted using biological replication, except for the virus-yield reduction assay (Fig. 1h), which relied on technical replication.

Research was conducted under an Institutional Animal Care and Use Committee approved protocol in compliance with the Animal Welfare Act, PHS Policy and other federal statutes and regulations relating to animals and experiments involving animals. The facility where this research was conducted is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Received 2 December 2012; accepted 13 January 2014.
Published online 2 March 2014.

1. Kuhn, J. H. *Filoviruses: A Compendium of 40 Years of Epidemiological, Clinical, and Laboratory Studies* (Springer, 2008).
2. Bausch, D. G. *et al.* Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. *J. Infect. Dis.* **196** (suppl. 2), S142–S147 (2007).
3. Kortepeter, M. G., Bausch, D. G. & Bray, M. Basic clinical and laboratory features of filoviral hemorrhagic fever. *J. Infect. Dis.* **204** (suppl. 3), S810–S816 (2011).
4. Towner, J. S. *et al.* Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* **4,** e1000212 (2008).
5. Geisbert, T. W. *et al.* Vesicular stomatitis virus-based vaccines protect nonhuman primates against aerosol challenge with Ebola and Marburg viruses. *Vaccine* **26,** 6894–6900 (2008).
6. Geisbert, T. W. *et al.* Postexposure protection of non-human primates against a lethal Ebola virus challenge with RNA interference: a proof-of-concept study. *Lancet* **375,** 1896–1905 (2010).
7. Warfield, K. L. *et al.* Ebola virus-like particle-based vaccine protects nonhuman primates against lethal Ebola virus challenge. *J. Infect. Dis.* **196** (suppl. 2), S430–S437 (2007).
8. Warren, T. K. *et al.* Advanced antisense therapies for postexposure protection against lethal filovirus infections. *Nature Med.* **16,** 991–994 (2010).
9. Dye, J. M. *et al.* Postexposure antibody prophylaxis protects nonhuman primates from filovirus disease. *Proc. Natl Acad. Sci. USA* **109,** 5034–5039 (2012).
10. Sullivan, N. J., Sanchez, A., Rollin, P. E., Yang, Z. Y. & Nabel, G. J. Development of a preventive vaccine for Ebola virus infection in primates. *Nature* **408,** 605–609 (2000).
11. Huggins, J. W. Prospects for treatment of viral hemorrhagic fevers with ribavirin, a broad-spectrum antiviral drug. *Rev. Infect. Dis.* **11** (suppl. 4), S750–S761 (1989).
12. Smither, S. J. *et al.* Post-exposure efficacy of Oral T-705 (Favipiravir) against inhalational Ebola virus infection in a mouse model. *Antiviral Res.* http://dx.doi.org/10.1016/j.antiviral.2014.01.012 (2014).
13. Aman, M. J. *et al.* Development of a broad-spectrum antiviral with activity against Ebola virus. *Antiviral Res.* **83,** 245–251 (2009).
14. Bray, M., Driscoll, J. & Huggins, J. W. Treatment of lethal Ebola virus infection in mice with a single dose of an *S*-adenosyl-L-homocysteine hydrolase inhibitor. *Antiviral Res.* **45,** 135–147 (2000).
15. Bray, M., Raymond, J. L., Geisbert, T. & Baker, R. O. 3-Deazaneplanocin A induces massively increased interferon-α production in Ebola virus-infected mice. *Antiviral Res.* **55,** 151–159 (2002).
16. Kinch, M. S. *et al.* FGI-104: a broad-spectrum small molecule inhibitor of viral infection. *Am. J. Transl. Res.* **1,** 87–98 (2009).
17. Panchal, R. G. *et al.* Identification of an antioxidant small-molecule with broad-spectrum antiviral activity. *Antiviral Res.* **93,** 23–29 (2012).
18. Warren, T. K. *et al.* Antiviral activity of a small-molecule inhibitor of filovirus infection. *Antimicrob. Agents Chemother.* **54,** 2152–2159 (2010).
19. Hevey, M., Negley, D., Pushko, P., Smith, J. & Schmaljohn, A. Marburg virus vaccines based upon alphavirus replicons protect guinea pigs and nonhuman primates. *Virology* **251,** 28–37 (1998).
20. Kobinger, G. P. *et al.* Replication, pathogenicity, shedding, and transmission of Zaire ebolavirus in pigs. *J. Infect. Dis.* **204,** 200–208 (2011).
21. Bente, D., Gren, J., Strong, J. E. & Feldmann, H. Disease modeling for Ebola and Marburg viruses. *Dis. Model. Mech.* **2,** 12–17 (2009).
22. Smith, D. H. *et al.* Marburg-virus disease in Kenya. *Lancet* **319,** 816–820 (1982).
23. Panchal, R. G. *et al.* Development of high-content imaging assays for lethal viral pathogens. *J. Biomol. Screen.* **15,** 755–765 (2010).
24. Barnard, D. L. *et al.* Inhibition of measles virus replication by 5′-nor carbocyclic adenosine analogues. *Antivir. Chem. Chemother.* **12,** 241–250 (2001).
25. Mühlberger, E., Weik, M., Volchkov, V. E., Klenk, H. D. & Becker, S. Comparison of the transcription and replication strategies of Marburg virus and Ebola virus by using artificial replication systems. *J. Virol.* **73,** 2333–2342 (1999).
26. Lou, H., Choi, Y. H., LaVoy, J. E., Major, M. E. & Hagedorn, C. H. Analysis of mutant NS5B proteins encoded by isolates from chimpanzees chronically infected following clonal HCV RNA inoculation. *Virology* **317,** 65–72 (2003).
27. Bray, M., Davis, K., Geisbert, T., Schmaljohn, C. & Huggins, J. A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J. Infect. Dis.* **178,** 651–661 (1998).
28. Warfield, K. L. *et al.* Development and characterization of a mouse model for Marburg hemorrhagic fever. *J. Virol.* **83,** 6404–6415 (2009).
29. Alves, D. A. *et al.* Aerosol exposure to the Angola strain of Marburg virus causes lethal viral hemorrhagic fever in cynomolgus macaques. *Vet. Pathol.* **47,** 831–851 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements J. Kuhn and J. Huggins provided insightful discussions and critically reviewed the manuscript. R. Kincaid and G. Feuerstein provided advice and guidance for BCX4430 development efforts. These studies were in part supported by The Joint Science and Technology Office for Chemical and Biological Defense of the Defense Threat Reduction Agency (proposal #TMTI0048_09_RD_T and CB3675) to S. Bavari. S. Radoshitzky assisted with the EBOV minigenome replicon assay. J. Reifman was essential to algorithm development of HCI image assessments. C. Basler provided the BHK-21-derived cell line constitutively expressing the T7 RNA polymerase. Plasmids encoding viral products and the EBOV minigenome replicon were provided by P. Kranzusch and S. Whelan. Neutral-red uptake antiviral assays were conducted by: D. L. Barnard, G. W. Day, B. Gowan, J. G. Julander, B. Tarbet, D. F. Smee and J. D. Morrey of Utah State University under National Institute of Allergy and Infectious Diseases (NIAID) contract HHSN272201100019I, BioQual Inc. under NIAID contract HHSN27220110005I, and at the University of Alabama Birmingham under NIAID contract HHSN272201100016I. Cell-based metabolism studies were conducted by C. Parker, X. Cheng, R. Upshaw and Y. Luo. A. Nalca, E. E. Zumbrun, H. Bloomfield, D. Dyer and J. Yeager assisted with virus aerosolization. C. Cooper provided assistance with the culture of human macrophage cell culture and R. Zamani provided assistance with high-content image assessments. S. Tritsch assisted with Nipah virus antiviral assays. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the US Army.

Author Contributions Y.S.B. and P.K. were responsible for the synthesis of BCX4430 and other small molecules. T.K.W. designed and supervised activities associated with rodent and non-human primate efficacy evaluations, evaluated study results, and wrote the manuscript. J.W., K.S.D., N.L.G. and S.A.V.T. conducted the rodent and non-human primate efficacy studies and performed sample analyses. R.G.P., G.P., C.J.R. and B.P.E. designed and executed cell-based filovirus assays and analysed these data. S. Bantia, Y.S.B., D.M.M., W.P.S., B.R.T. and others designed and analysed data from cell-based antiviral assays. L.D. conducted quantitative PCR analysis. B.R.T. conducted statistical evaluations of *in vivo* study results. S.H. performed post-mortem analysis of all non-human primate subjects. Y.S.B. supervised the pharmacokinetics studies of BCX4430 and W.P.S. conducted pharmacokinetics data analysis. S. Bantia conducted assessments of BCX4430 metabolite analysis and incorporation into host nucleic acids. X.C. conducted chain termination experiments. T.K.W., D.M.M., L.S.W., B.R.T., Y.S.B., W.P.S. and S. Bavari designed experiments, evaluated results and provided project oversight.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Bavari (sina.bavari.civ@mail.mil).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13204

Caenorhabditis elegans pathways that surveil and defend mitochondria

Ying Liu[1,2]†, Buck S. Samuel[1,2], Peter C. Breen[1,2] & Gary Ruvkun[1,2]

Mitochondrial function is challenged by toxic by-products of metabolism as well as by pathogen attack[1,2]. *Caenorhabditis elegans* normally responds to mitochondrial dysfunction with activation of mitochondrial-repair, drug-detoxification and pathogen-response pathways[1–7]. Here, from a genome-wide RNA interference (RNAi) screen, we identified 45 *C. elegans* genes that are required to upregulate detoxification, pathogen-response and mitochondrial-repair pathways after inhibition of mitochondrial function by drug-induced or genetic disruption. Animals defective in ceramide biosynthesis are deficient in mitochondrial surveillance, and addition of particular ceramides can rescue the surveillance defects. Ceramide can also rescue the mitochondrial surveillance defects of other gene inactivations, mapping these gene activities upstream of ceramide. Inhibition of the mevalonate pathway, either by RNAi or statin drugs, also disrupts mitochondrial surveillance. Growth of *C. elegans* with a significant fraction of bacterial species from their natural habitat causes mitochondrial dysfunction. Other bacterial species inhibit *C. elegans* defence responses to a mitochondrial toxin, revealing bacterial countermeasures to animal defence.

Mitochondria are almost entirely composed of proteins encoded by the nuclear genome[8]. A surveillance pathway detects mitochondrial defects to induce mitochondrial chaperone genes *hsp-6* and *hsp-60* and xenobiotic-detoxification and pathogen-response pathways[1–7]. Disrupting *C. elegans* mitochondrial function induces drug-detoxification genes such as *cyp-14A3* and *ugt-61* (Fig. 1a–c), as well as a *Pseudomonas* pathogen-response gene *irg-1* (ref. 9) (Fig. 1d). This suggests that animals interpret a disruption of mitochondrial activity as a xenobiotic or pathogen attack. Consistent with this, *C. elegans* exhibits bacterial-avoidance behaviour when mitochondria are inhibited by RNAi or antimycin, an inhibitor of mitochondrial electron transport produced by *Streptomyces*[1] (Fig. 1e, f). Animals also show aversive behaviours in the absence of any drug or mitochondrial gene inactivation if they carry a mutation of the nuclearly encoded mitochondrial gene *isp-1(qm150)* (Fig. 1g). This mutation also activates *hsp-6* (Fig. 1h).

C. elegans live in microbe-rich environments[10]. We isolated microbes from habitats with wild *C. elegans* populations[10], classified them based on their 16S ribosomal sequences (B.S.S., M. A. Félix & G.R., unpublished observations), and found that 18% of the 560 strains tested caused mitochondrial stress as revealed by induction of *hsp-6p::gfp* (where *p* indicates promoter and *gfp* indicates the gene encoding green fluorescent protein) (Fig. 2a–c). This suggests that mitochondria are targeted by many bacterial species, and explains the evolutionary advantage of coupling detection of *C. elegans* mitochondrial dysfunction to antibacterial gene expression and behavioural responses. Iron is a valuable resource to microbes, which often produce siderophores to capture the iron; the eukaryotic mitochondrion, rich in haem and iron sulphur proteins,

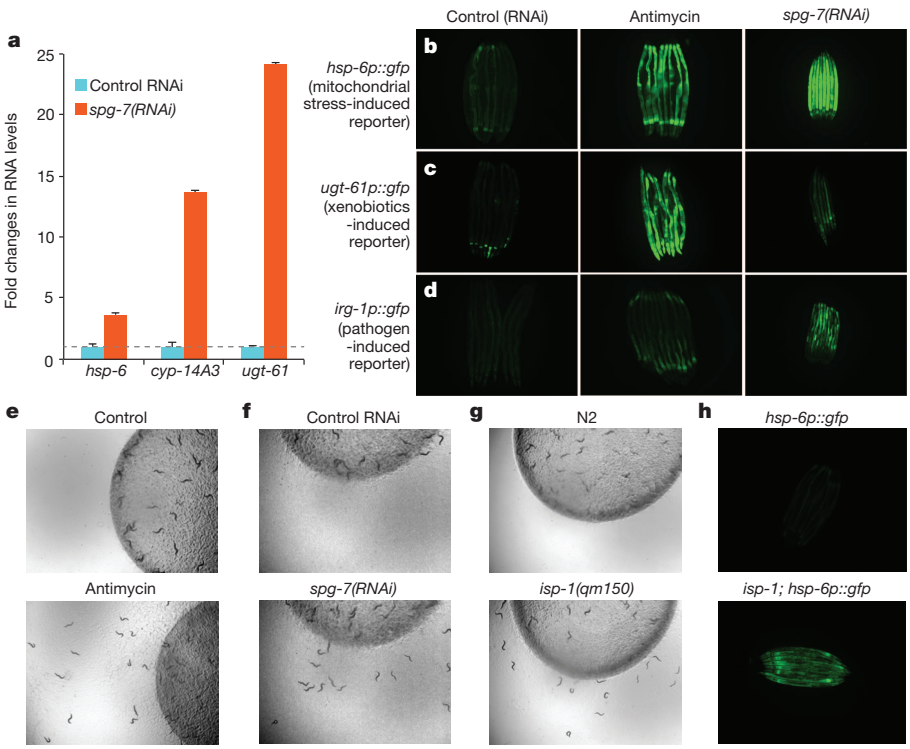


Figure 1 | Mitochondrial dysfunction activates homeostatic, detoxification and pathogen responses. **a**, Mitochondrial chaperone and detoxification transcripts in control versus *spg-7(RNAi)* ($n = 3$). Error bars represent s.d. **b–d**, *hsp-6p::gfp* (**b**), *ugt-61p::gfp* (**c**) and *irg-1p::gfp* (**d**) animals raised on control *Escherichia coli*, or *E. coli* with antimycin or *spg-7(RNAi)*. **e–g**, Drug (**e**)-, RNAi (**f**)- or mutant allele (**g**)-based mitochondrial inhibition causes food avoidance. N2, wild type. **h**, Fluorescence images of wild-type *hsp-6p::gfp* and *isp-1(qm150);hsp-6p::gfp* animals.

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present address: State Key Laboratory of Biomembrane and Membrane Biotechnology, Institute of Molecular Medicine, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

©2014 Macmillan Publishers Limited. All rights reserved

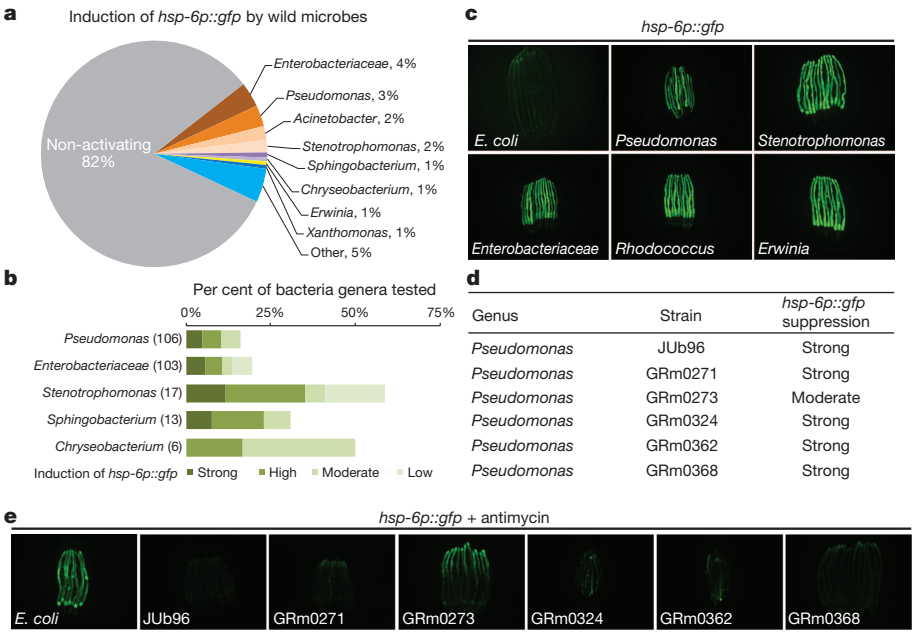


Figure 2 | Some bacterial species from the *C. elegans* natural habitat antagonize the mitochondria. **a**, Genera that induce *hsp-6p::gfp* (100 of 560 tested microbes (18%)). **b**, Proportion of bacterial genera tested that cause mitochondrial dysfunction and *hsp-6p::gfp* induction. **c**, *hsp-6p::gfp* animals raised on *E. coli* or natural microbial species. **d**, Six bacterial strains that render *C. elegans* defective in *hsp-6* response to antimycin. **e**, *hsp-6p::gfp* animals exposed to antimycin and raised on control *E. coli* or the six microbial strains.

is an attractive iron depot for bacteria. Bacterial toxins, siderophores or virulence factors from these many bacterial strains may target the mitochondrion.

To identify the molecular basis behind the detection of animal mitochondrial dysfunction and its coupling to a detoxification response, we performed a genome-wide RNAi screen for gene inactivations that render *C. elegans* unresponsive to mitochondrial dysfunction (Extended Data Figs 1 and 2). Gene inactivations that caused a failure to induce *hsp-6p::gfp* in antimycin were re-screened for failure to upregulate other genes normally induced by mitochondrial dysfunction: the detoxification gene *ugt-61::gfp* in the presence of antimycin, *hsp-6p::gfp* induction in the *isp-1(qm150)* mutant, or *hsp-6p::gfp* induction in response to RNAi targeting the ATP synthase *atp-2*. Mitochondrial surveillance and response was disrupted by 45 gene inactivations (Extended Data Table 1), including the known components *atfs-1*, *clpp-1* and *dve-1* (refs 4, 6).

One of the gene inactivations that potently disrupts response to mitochondrial dysfunction, *sptl-1*, encodes serine palmitoyltransferase, which has a role in the sphingolipid biosynthesis pathway (Extended Data Fig. 3). Inactivation of *sptl-1* by RNAi inhibited the induction of *hsp-6p::gfp* upon antimycin treatment or *spg-7(RNAi)* (Fig. 3a, b), whereas *sptl-1(RNAi)* alone without antimycin treatment did not induce the expression of *hsp-6p::gfp* (Extended Data Fig. 4b). *sptl-1(RNAi)* did not affect the activation of endoplasmic reticulum stress reporter *hsp-4p::gfp* by the drug tunicamycin (Fig. 3c), indicating a specific role in mitochondrial surveillance. A probable null mutation in *sptl-1* (Extended Data Fig. 5a) impaired the induction of *hsp-6*, *cyp-14A3* and *ugt-61* after mitochondrial disruption by *spg-7(RNAi)* (Fig. 3d). A double mutant in the two ceramide synthase genes of the sphingolipid biosynthetic pathway also attenuated the induction of *hsp-6* upon mitochondrial damage (Extended Data Fig. 5b). Treatment with myriocin, a fungal inhibitor of mammalian serine palmitoyltransferase, disrupted antimycin-induced *hsp-6p::gfp* induction (Fig. 3e). *sptl-1(RNAi)* (with normal movement, Supplementary Videos 1 and 2) attenuated food avoidance induced by antimycin or *spg-7(RNAi)*, suggesting that this behavioural response is also coupled to *sptl-1* (Fig. 3g and Extended Data Fig. 4d, e). *sptl-1* expression is upregulated 2.5-fold by mitochondrial damage (Extended Data Fig. 4a), suggesting that increased sphingolipid levels during mitochondrial disruption may act as a signal rather than as a membrane component required for another signal.

The morphology of the normally extensive mitochondrial network of *C. elegans* and other animals is responsive to mutation or inactivation of mitochondrial components[11]. Mitochondria in *sptl-1(RNAi)* animals hyperfused into larger, longer mitochondria (Fig. 3f), suggesting that *sptl-1* is critical for mitochondrial homeostasis. The *isp-1* mitochondrial mutant had a severe synthetic growth defect on a *sptl-1(RNAi)* background, whereas wild-type development was only slightly delayed on *sptl-1(RNAi)* (Fig. 3h). Sphingolipid signalling may act in the homeostatic response to the mitochondrial defect caused by the *isp-1(qm150)* mutation; in the absence of that response, *isp-1(qm150)* may cause a more severe mitochondrial defect and growth arrest. *sptl-1(RNAi)* animals were also more sensitive to the mitochondrial inhibitor antimycin (Extended Data Fig. 4f). In addition, *sptl-1(RNAi)* animals were unable to sense mitochondrial inhibition and activate the mitochondrial surveillance pathway when challenged by a set of the wild microbes that disrupt mitochondrial function (Extended Data Fig. 4g).

Ceramide supplementation rescued the *sptl-1* defect in mitochondrial surveillance (Fig. 4a and Extended Data Fig. 5d), whereas ceramide in the absence of mitochondrial damage did not induce *hsp-6p::gfp* (Extended Data Fig. 5e), suggesting that ceramide is not a sufficient signal to induce the suite of responses to mitochondrial dysfunction in the absence of true mitochondrial dysfunction. Ceramide supplementation also partially rescued the mitochondrial morphology defect (Fig. 4b) and the impaired food-avoidance behaviour caused by *sptl-1(RNAi)* (Extended Data Fig. 6a). Dihydroceramide, a ceramide precursor with no signalling function in mammals, could not rescue the *sptl-1(RNAi)* mitochondrial surveillance defect (Fig. 4a). Two *C. elegans* ceramide synthases, HYL-1 and HYL-2, synthesize different ceramide species (C24 to C26 and C20 to C22, respectively[12]). On testing C16, C20, C22 and C24 ceramides, only C24 rescued the deficiency of mitochondrial surveillance caused by *sptl-1(RNAi)* (Fig. 4c).

During apoptosis and mitophagy, ceramide accumulates on the outer membrane of mitochondria[13–15]. Using antibodies to stain ceramide and Complex IV (COX-IV; a mitochondrial protein in the electron transport chain), ceramide and mitochondrial protein colocalization was shown to markedly increase after *C. elegans* mitochondrial inhibition (Fig. 4d).

©2014 Macmillan Publishers Limited. All rights reserved

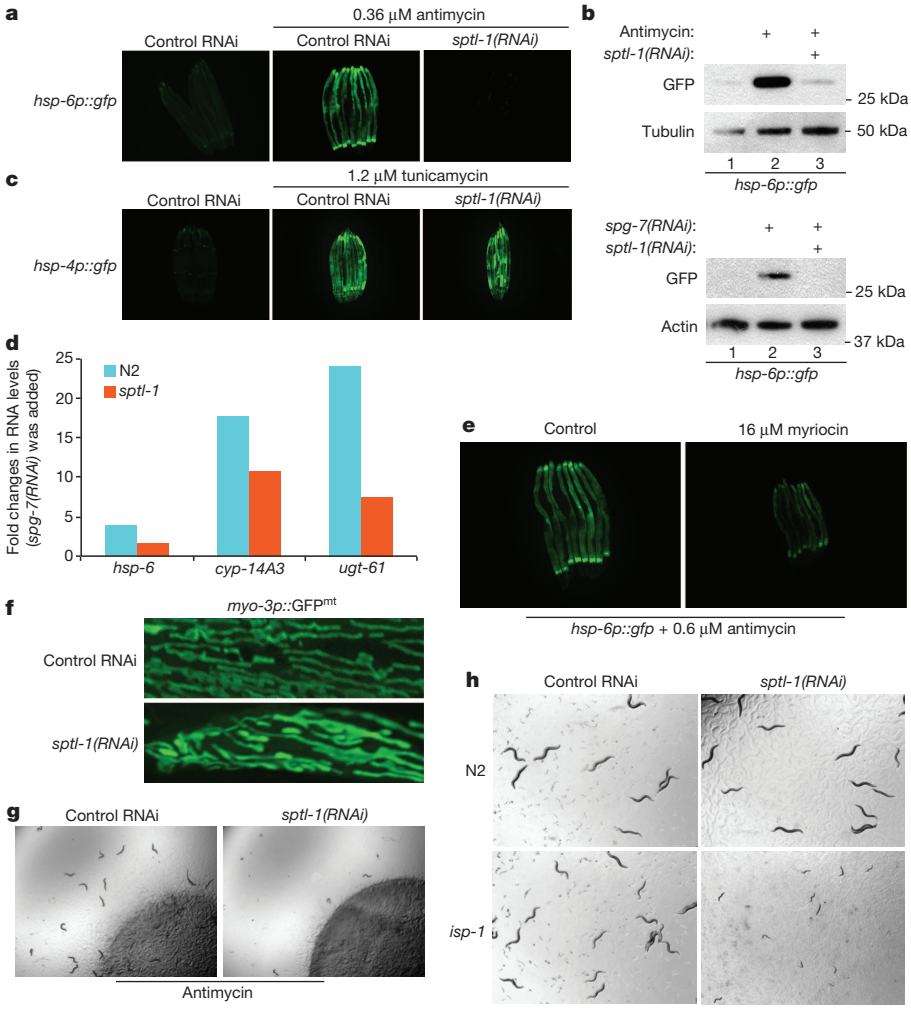


Figure 3 | The serine palmitoyltransferase *sptl-1* is required for mitochondrial surveillance. **a**, *hsp-6p::gfp* animals raised on control or *sptl-1(RNAi)* in the presence or absence of antimycin. **b**, GFP immunoblot expressed by *hsp-6p::gfp* animals in the presence or absence of antimycin or *spg-7(RNAi)*. **c**, *hsp-4p::gfp* animals raised on control or *sptl-1(RNAi)* in the presence or absence of the endoplasmic reticulum drug tunicamycin. **d**, *C. elegans hsp-6* and xenobiotic-detoxification gene transcripts in control or *sptl-1(ok1693)* mutant animals after exposure to *spg-7(RNAi)* ($n = 2$). **e**, *hsp-6p::gfp* animals raised in the presence or absence of myriocin. **f**, Body wall muscle of *sptl-1(RNAi)* compared to wild-type animals expressing a mitochondrially localized GFP reporter. **g**, Food-avoidance phenotypes for control or *sptl-1(RNAi)* animals treated with antimycin. **h**, Wild type (N2) or *isp-1(qm150)* raised on control or *sptl-1(RNAi)*.

The increased colocalization preceded induction of *hsp-6p::gfp* (Extended Data Fig. 6b, c). Thus, ceramide may participate in an early step of mitochondrial surveillance by marking domains of dysfunction.

To map ceramide relative to other gene inactivations that render animals unable to respond to mitochondrial damage, we tested whether ceramide could rescue the mitochondrial surveillance defects of other hits from the RNAi screen. Five other gene inactivations were rescued by ceramide: *ran-4*, a nuclear transport component; *Y47G6A.29*, a phosphatidylinositol signalling component; *F40F12.7*, a zinc finger protein; *Y54E10BR.5*, a signal peptidase component; and *ceh-20*, a homeobox transcription factor (Extended Data Fig. 6e). These gene inactivations may disrupt mitochondrial surveillance upstream of ceramide production.

ATFS-1 is a transcription factor that activates *hsp-6* and *hsp-60* during mitochondrial stress (Fig. 4e)[5]. ATFS-1 contains an amino-terminal mitochondrial targeting sequence as well as a carboxy-terminal nuclear localization signal; when mitochondria are damaged, nuclear accumulation of ATFS-1 is favoured[6]. Deletion of the N-terminal 1–32 amino acids causes constitutive nuclear accumulation of ATFS-1 (Extended Data Fig. 6d) and activates *hsp-60p::gfp*[6]. Activation of *hsp-60p::gfp* by ATFS-1$^{\Delta 1\text{-}32.myc}$ did not require *sptl-1* gene activity (Fig. 4f), suggesting that ceramide works upstream of ATFS-1 and has a role in the early detection of mitochondrial dysfunction.

hmgs-1, which encodes an HMG-CoA (3-hydroxy-3-methyl-glutaryl-coenzyme A) synthase of the mevalonate synthetic pathway, is also a strong hit from the genome-wide screen. *hmgs-1* gene inactivation inhibited antimycin-induced *hsp-6p::gfp* induction and food avoidance (Extended Data Fig. 7a–d). *hmgs-1(RNAi)* also induced abnormal mitochondrial morphology (Extended Data Fig. 7e). Supplementing mevalonate to *hmgs-1(RNAi)* animals rescued the deficiency of antimycin-induced *hsp-6p::gfp* (Extended Data Fig. 7f).

Statins are cholesterol-lowering drugs that inhibit the mevalonate pathway (Extended Data Fig. 8a)[16]. Treating *C. elegans* with statins also abrogates their ability to sense mitochondrial damage and activate protective programs (Extended Data Figs 7g and 8b). A common side effect of statin therapy is muscle toxicity[17]. By inhibiting the mevalonate pathway, statins inhibit ubiquinone synthesis, a key component of the electron transport chain (Extended Data Fig. 8a). Our data suggest that a combined inhibition of ubiquinone synthesis and mitochondrial surveillance may contribute to the muscle toxicity of statins. In support of a role for mitochondrial surveillance in statin action, a gain-of-function mutation in *C. elegans atfs-1* confers statin resistance[18]. Treating human embryonic kidney HEK293T cells with statin impaired the mitochondrial network and morphology (Extended Data Figs 7h and 9). Statin treatment also decreased ATP production in mouse C2C12 myotubes (Extended Data Fig. 7i).

Eukaryotic mitochondrial surveillance pathway components are likely to be targets of microbial toxins and virulence factors. If the mitochondrial surveillance pathway in an animal is disabled by a bacterial toxin or virulence factor, other anti-mitochondrial toxins, siderophores or virulence factors would be rendered more effective. To detect such bacterial anti-surveillance activities, we screened our collection of *C. elegans* flora for bacterial species that, when co-cultured with *C. elegans*, could disrupt the induction of *hsp-6p::gfp* by antimycin. Six wild bacterial strains of the genus *Pseudomonas* from three species (*vranovensis*, *brenneri*

©2014 Macmillan Publishers Limited. All rights reserved

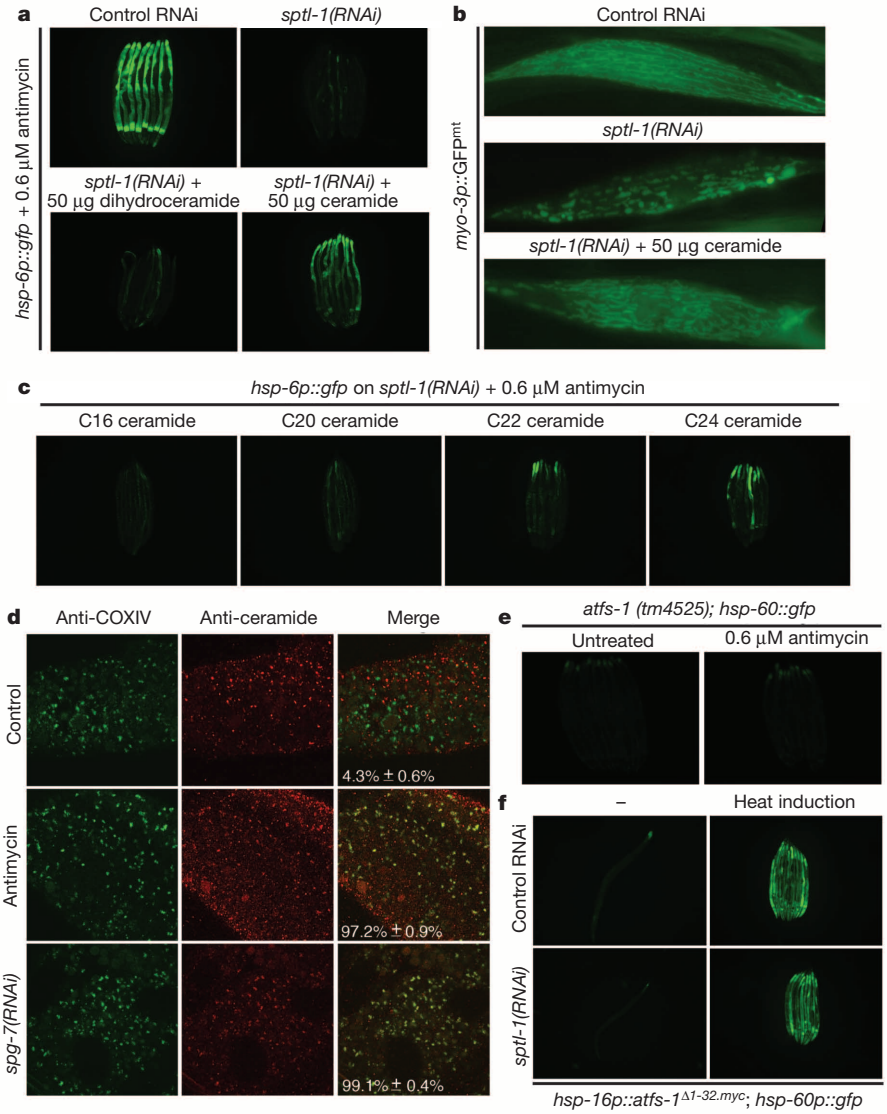


Figure 4 | Ceramide biogenesis is required for mitochondrial surveillance. **a**, *hsp-6p::gfp* animals exposed to antimycin in the presence or absence of dihydroceramide or ceramide. **b**, Body wall muscle of animals expressing a mitochondrially localized GFP reporter in *sptl-1(RNAi)* with and without added ceramide. **c**, *hsp-6p::gfp sptl-1(RNAi)* plus antimycin with different ceramide species. **d**, Dissected animals after antimycin or *spg-7(RNAi)* treatment were stained with anti-COX-IV (green) and anti-ceramide antibodies (red). Quantification represents proportion of mitochondria with contact of mitochondrial and ceramide staining (mean ± s.d., $n = 3$). **e**, *atfs-1(tm4525);hsp-60p::gfp* animals alone or with antimycin. **f**, *atfs-1(tm4525);hsp-16p::atfs-1$^{\Delta 1\text{-}32.myc}$;hsp-60p::gfp* animals raised on control or *sptl-1(RNAi)*.

and *asplenii*) disrupt mitochondrial surveillance (Fig. 2d, e). The dozens of *C. elegans* genes we have identified in the mitochondrial surveillance pathway are candidate targets for toxins or virulence factors from these *Pseudomonas* strains.

Our studies have revealed roles of ceramide and mevalonate in mitochondrial surveillance. The products of these pathways either constitute signals that are transferred within or between cells, or structures within those cells that are necessary for other signals to emerge. We favour that these constitute signals because the expression of these biosynthetic pathways are induced by the mitochondrial insults (Extended Data Fig. 4a) and because these molecules are localized to the site of injury (Fig. 4d). Ceramides are upstream elements of the stress response of plasma membrane, endoplasmic reticulum and mitochondria in multiple species[19,20], and can act as hormonal signals[21]. But it is also possible that these constitute membrane elements required for other signals. Variations in the animal pathways for mitochondrial surveillance identified here, many of which involve conserved proteins, may underlie variation in the symptoms caused by the same mitochondrial mutations or variation in response to mitochondrial toxins such as statins. Microbial secondary metabolites that render animals unresponsive to their mitochondrial dysfunctions (Fig. 2d, e) may have therapeutic potential in the treatment of a predicted upregulation of detoxification and antibacterial pathways, which may contribute to the devastating symptoms of some mitochondrial diseases. The anti-surveillance activities of such natural products may also treat other disorders of dysregulated detoxification and innate immunity, such as autoimmunity.

METHODS SUMMARY

RNAi clones in *E. coli* were seeded to 24-well worm plates and ~40 worms per well were assayed at 48 h. For drug screens, GFP expression was assayed after 24 h. Wild microbes were seeded onto 24-well plates in duplicate and ~40 synchronized L1 worms were added to the wells. Animals were scored after 48 h.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 June 2013; accepted 4 March 2014.
Published online 2 April 2014.

1. Melo, J. A. & Ruvkun, G. Inactivation of conserved *C. elegans* genes engages pathogen- and xenobiotic-associated defenses. *Cell* **149,** 452–466 (2012).
2. Runkel, E. D., Liu, S., Baumeister, R. & Schulze, E. Surveillance-activated defenses block the ROS-induced mitochondrial unfolded protein response. *PLoS Genet.* **9,** e1003346 (2013).
3. Benedetti, C., Haynes, C. M., Yang, Y., Harding, H. P. & Ron, D. Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response. *Genetics* **174,** 229–239 (2006).

©2014 Macmillan Publishers Limited. All rights reserved

4. Haynes, C. M., Petrova, K., Benedetti, C., Yang, Y. & Ron, D. ClpP mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Dev. Cell* **13,** 467–480 (2007).
5. Haynes, C. M., Yang, Y., Blais, S. P., Neubert, T. A. & Ron, D. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* **37,** 529–540 (2010).
6. Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation. *Science* **337,** 587–590 (2012).
7. Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144,** 79–91 (2011).
8. Pellegrino, M. W., Nargund, A. M. & Haynes, C. M. Signaling the mitochondrial unfolded protein response. *Biochim. Biophys. Acta* **1833,** 410–416 (2013).
9. Estes, K. A., Dunbar, T. L., Powell, J. R., Ausubel, F. M. & Troemel, E. R. bZIP transcription factor *zip-2* mediates an early response to *Pseudomonas aeruginosa* infection in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **107,** 2153–2158 (2010).
10. Félix, M. A. & Duveau, F. Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. *BMC Biol.* **10,** 59 (2012).
11. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33,** 40–48 (2003).
12. Menuz, V. *et al.* Protection of *C. elegans* from anoxia by HYL-2 ceramide synthase. *Science* **324,** 381–384 (2009).
13. Deng, X. *et al.* Ceramide biogenesis is required for radiation-induced apoptosis in the germ line of *C. elegans*. *Science* **322,** 110–115 (2008).
14. Sentelle, R. D. *et al.* Ceramide targets autophagosomes to mitochondria and induces lethal mitophagy. *Nature Chem. Biol.* **8,** 831–838 (2012).
15. Colombini, M. Ceramide channels and their role in mitochondria-mediated apoptosis. *Biochim. Biophys. Acta* **1797,** 1239–1244 (2010).
16. Wenner Moyer, M. The search beyond statins. *Nature Med.* **16,** 150–153 (2010).
17. Graham, D. J. *et al.* Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *J. Am. Med. Assoc.* **292,** 2585–2590 (2004).
18. Rauthan, M., Ranji, P., Aguilera Pradenas, N., Pitot, C. & Pilon, M. The mitochondrial unfolded protein response activator ATFS-1 protects cells from inhibition of the mevalonate pathway. *Proc. Natl Acad. Sci. USA* **110,** 5981–5986 (2013).
19. Hage-Sleiman, R., Esmerian, M. O., Kobeissy, H. & Dbaibo, G. p53 and ceramide as collaborators in the stress response. *Int. J. Mol. Sci.* **14,** 4982–5012 (2013).
20. Senkal, C. E. *et al.* Alteration of ceramide synthase 6/C16-ceramide induces activating transcription factor 6-mediated endoplasmic reticulum (ER) stress and apoptosis via perturbation of cellular Ca^{2+} and ER/Golgi membrane network. *J. Biol. Chem.* **286,** 42446–42458 (2011).
21. Breslow, D. K. & Weissman, J. S. Membranes in balance: mechanisms of sphingolipid homeostasis. *Mol. Cell* **40,** 267–279 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Haynes (Memorial Sloan Kettering Cancer Center), F. Ausubel (Massachusetts General Hospital) and the Caenorhabditis Genetics Center for providing strains; and the V. Mootha laboratory (Massachusetts General Hospital) for providing C2C12 cells. We thank J. Larkins-Ford, C. Phillips, S. Garcia and M. Pellegrino for technical support, as well as M. A. Félix for bacterial strains and for guidance in the collection of microbial strains from *C. elegans* natural habitats. Y.L. is supported by the Helen Hay Whitney research fellowship; B.S.S. is supported by the Charles King Trust postdoctoral fellowship. The work is supported by a grant from the National Institutes of Health, awarded to G.R. (NIH AG043184-16).

Author Contributions Y.L. and G.R. designed experiments; Y.L., B.S.S. and P.C.B. carried out experiments. Y.L., B.S.S. and G.R. wrote the paper. G.R. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.R. (ruvkun@molbio.mgh.harvard.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13069

miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*

Kate M. Creasey[1], Jixian Zhai[2], Filipe Borges[1], Frederic Van Ex[1], Michael Regulski[1], Blake C. Meyers[2] & Robert A. Martienssen[1,3,4]

In plants, post-transcriptional gene silencing (PTGS) is mediated by DICER-LIKE 1 (*DCL1*)-dependent microRNAs (miRNAs), which also trigger 21-nucleotide secondary short interfering RNAs (siRNAs) via RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), DCL4 and ARGONAUTE 1 (AGO1)[1–3], whereas transcriptional gene silencing (TGS) of transposons is mediated by 24-nucleotide heterochromatic (het)siRNAs, *RDR2, DCL3* and *AGO4* (ref. 4). Transposons can also give rise to abundant 21-nucleotide 'epigenetically activated' small interfering RNAs (easiRNAs) in *DECREASED DNA METHYLATION 1* (*ddm1*) and *DNA METHYLTRANSFERASE 1* (*met1*) mutants, as well as in the vegetative nucleus of pollen grains[5] and in dedifferentiated plant cell cultures[6]. Here we show that easiRNAs in *Arabidopsis thaliana* resemble secondary siRNAs, in that thousands of transposon transcripts are specifically targeted by more than 50 miRNAs for cleavage and processing by *RDR6*. Loss of *RDR6, DCL4* or *DCL1* in a *ddm1* background results in loss of 21-nucleotide easiRNAs and severe infertility, but 24-nucleotide hetsiRNAs are partially restored, supporting an antagonistic relationship between PTGS and TGS. Thus miRNA-directed easiRNA biogenesis is a latent mechanism that specifically targets transposon transcripts, but only when they are epigenetically reactivated during reprogramming of the germ line. This ancient recognition mechanism may have been retained both by transposons to evade long-term heterochromatic silencing and by their hosts for genome defence.

Twenty-one-nucleotide epigenetically activated small RNAs, which we have termed easiRNAs, were previously found to accumulate from the 3′ untranslated region (UTR) of *ATHILA6* retrotransposons in wild-type pollen and *ddm1* inflorescence[5], and to depend on *RDR6*, *DCL4* and *AGO1* (refs 7, 8). To assess the origin of easiRNAs, we performed small RNA sequencing from inflorescence tissue, and we found that they accumulated in *ddm1*, but not in wild type (Columbia-0 (Col-0)) nor in *ddm1 rdr6* double mutants[7,8] (Fig. 1 and Extended Data Fig. 1a). Twenty-one-nucleotide easiRNAs in *ddm1* mostly originated from *ATGYPSY* long terminal repeat (LTR) retroelements (Supplementary Table 1) located in pericentromeric regions, especially the high copy but defective *ATHILA* retrotransposons, which are integrated into pericentromeric satellite repeats (Fig. 1). However, easiRNAs also arose from *ATCOPIA* families found in euchromatic regions, such as *ATCOPIA93* (also known as *EVADÉ*)[9] (Supplementary Table 1), and from specific *VANDAL* (also known as *MuDR*), *HAT* and *CACTA* elements (such as *ATENSPM6*), some of which are present in much lower copy numbers and known to transpose in *ddm1* (ref. 10). This raised the important question as to how specificity was conferred to these transposable elements (TEs) such that they gave rise to easiRNAs, when other transposons and genes did not. Genetically, easiRNAs resemble 21-nucleotide *trans*-acting (ta)siRNAs and other secondary siRNAs, which are derived from non-coding RNAs and messenger RNAs, respectively, and are triggered by miRNAs bound by *AGO7* (ref. 11) and *AGO1* (refs 2, 3, 12, 13). miRNAs were thus excellent candidates to confer transposon specificity.

miRNA biogenesis in *Arabidopsis* depends largely on *DCL1*, such that miRNAs are greatly reduced in sterile *dcl1-9* plants and partly reduced in the weaker fertile allele *dcl1-11* (ref. 14). We were unable to recover *ddm1 dcl1*-9 double mutants, but easiRNA levels were reduced in *ddm1 dcl1-11*, consistent with a role for miRNA in targeting and easiRNA biogenesis (Fig. 1 and Extended Data Fig. 1b). We identified miRNAs in our small RNA sequencing libraries from Col-0, *ddm1* and *ddm1 rdr6* inflorescence tissue. In addition to using a miRNA identification algorithm[15], miRNAs were distinguishable from other 21-nucleotide sRNAs that are *RDR6* dependent by comparing their abundance in *ddm1* and *ddm1 rdr6* small RNA sequencing libraries. We identified several known miRNAs, validating our algorithm (Extended Data Fig. 10 and Supplementary Table 2), and genome-wide target prediction revealed that 3,662 TEs are potentially targeted by these miRNAs (Supplementary Table 3). To determine whether miRNAs mediate targeting and cleavage of TE transcripts, we sequenced cleavage sites genome wide by parallel analysis of RNA ends (PARE)[11]. To assess specificity, 5′ RNA ends from a 5-nucleotide window surrounding each predicted miRNA target site were compared to those from a larger 30-nucleotide window (Supplementary Table 3; for more specific 1-nucleotide and 3-nucleotide windows, refer to Supplementary Table 9). Significant enrichment of RNA ends at or near the target site was strong evidence for cleavage of the transposon transcripts guided by miRNA.

Approximately half of the 3,662 predicted TE targets showed evidence of miRNA-guided cleavage (Supplementary Table 3), and were targeted by more than 50 distinct miRNAs (Extended Data Fig. 10), although some TEs had only one cleavage product in the target window (such as *ATCOPIA93*, which is predicted to be cleaved by miR-833). In some cases multiple miRNAs targeted single TEs (Extended Data Fig. 10), reminiscent of the 'two-hit' model for tasiRNA biogenesis[16]. In other cases, TEs were targeted by longer forms of miRNA (22 nucleotides), which are thought to promote secondary siRNA biogenesis[3,17]. Specifically, 1,733 TEs were predicted to be miRNA targets and generated easiRNAs (at least ten reads). Of these, 1,247 were detectably cleaved by PARE sequencing. An additional 1,929 TEs were predicted to be targeted by miRNAs but did not generate easiRNAs, and of these 442 were detectably cleaved by PARE-seq. Thus, more than half of the TEs targeted by miRNAs also generated easiRNAs in inflorescence tissue from *ddm1*. However, phasing of easiRNAs, typical of other secondary siRNAs, was not detected, probably due to the repetitive nature of the TE targets and their targeting by multiple miRNAs (data not shown).

Interestingly, the miRNAs found to target TEs were mostly known miRNAs (Extended Data Fig. 10), such as miR-156, miR-159, miR-172 and miR-859, which also generate secondary siRNAs from mRNA targets[2,3]. Many of these miRNAs were predicted to target *ATHILA* elements (Extended Data Fig. 1), generating abundant easiRNA corresponding to *ATHILA OPEN READING FRAME 1* (*ORF1*), also known as *TRANSCRIPTIONALLY SILENT INFORMATION* (*TSI*)

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Delaware Biotechnology Institute and Department of Plant & Soil Sciences, 15 Innovation Way, University of Delaware, Newark, Delaware 19711, USA. [3]Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, New York 11724, USA. [4]Chaire Blaise Pascal, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), 75230 Paris, France.

©2014 Macmillan Publishers Limited. All rights reserved

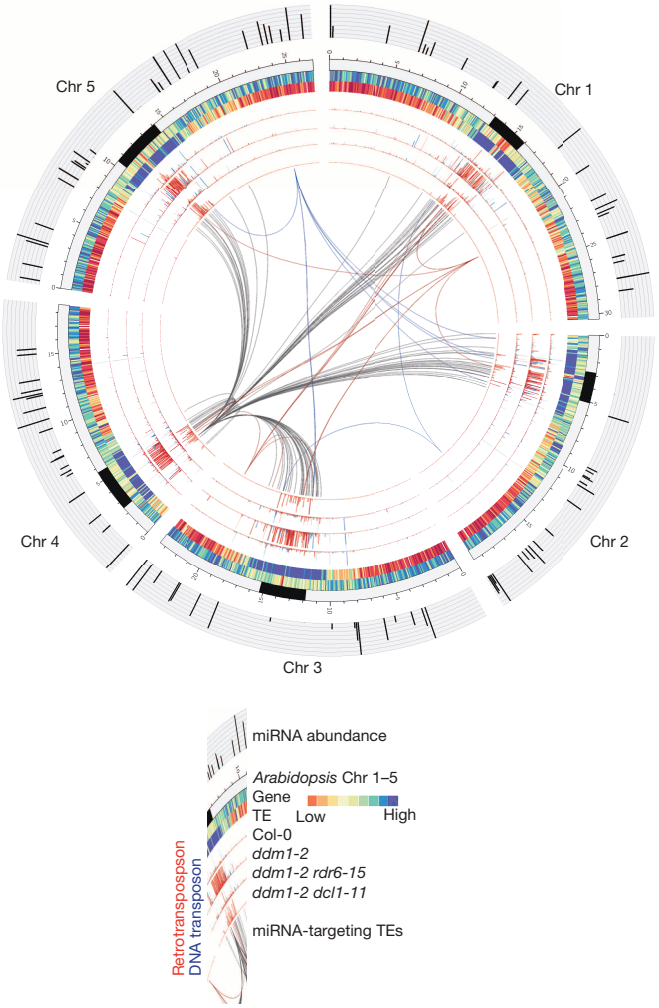


Figure 1 | miRNAs trigger *RDR6*-dependent 21-nucleotide easiRNA biogenesis from reactivated transposons. Whole-genome representation illustrating miRNAs triggering widespread easiRNA biogenesis at transposons in *Arabidopsis*. Outermost to innermost tracks depicting: miRNA abundance in *ddm1-2* (histogram); *Arabidopsis* chromosome (Chr) 1–5 pericentromeric region (black) (ideogram); transposon density (heat map: low density, blue; high density, red); gene density (heat map); retrotransposon-derived 21-nucleotide easiRNAs (dark red, unique; light red, multiple mapping) and DNA transposon-derived 21-nucleotide small RNAs (dark blue, unique; light blue, multiple mapping) in the order Col-0, *ddm1-2*, *ddm1-2 rdr6-15* and *ddm1-2 dcl1-11* (histograms); miRNAs targeting transposons (connectors); miR-859a (Chr 1, red), miR-390a (Chr 2, blue), miR-172d (Chr 3, red), eamiR-2 from *ATHILAIV* (Chr 4, grey) and miR-172e (Chr 5, blue). Transposons are post-transcriptionally targeted by 50 known miRNAs and newly discovered eamiRNAs (Extended Data Fig. 10), giving rise to abundant *RDR6*-dependent 21-nucleotide easiRNAs from transposons in *Arabidopsis*. (See also Extended Data Fig. 1 and Supplementary Tables 1, 3.)

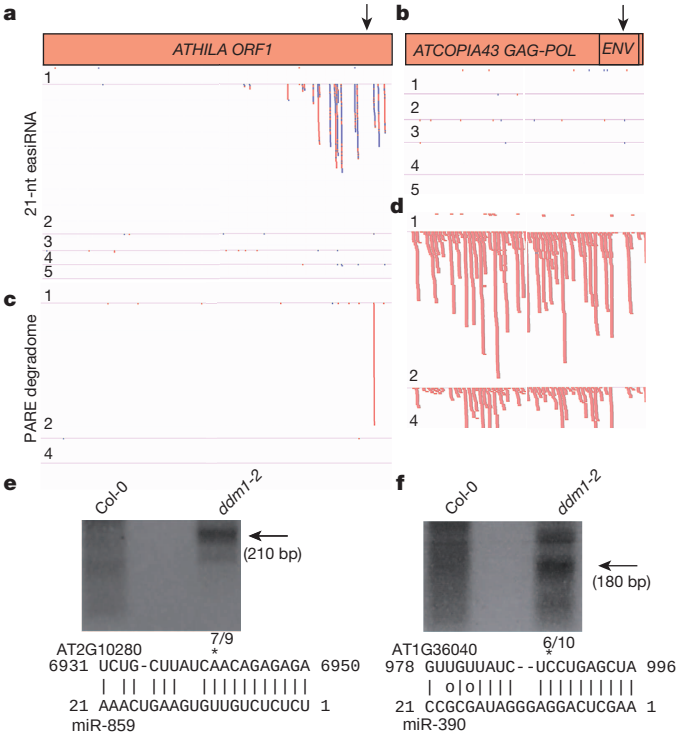


Figure 2 | miRNA cleavage at *ATHILA ORF1* and *ATCOPIA43*. **a**, **b**, Twenty-one-nucleotide (nt) easiRNAs at *ATHILA ORF1* (The *Arabidopsis* Information Resource (TAIR) accession AT2G10280), in comparison with non-easiRNA generating *ATCOPIA43* (AT1G36040). Track order: 1, Col-0; 2, *ddm1-2*; 3, *rdr6-15*; 4, *ddm1-2 rdr6-15*; 5, *ddm1-2 dcl1-11*. **c**, **d**, PARE at *ATHILA ORF1* and *ATCOPIA43*. Track order: 1, Col-0; 2, *ddm1-2*; 4, *ddm1-2 rdr6-15*. Individual reads: sense, red; antisense, blue. **e**, **f**, 5′ RLM RACE PCR products (arrows indicate black bars in **a**, **b**) from *ATHILA ORF1* and *ATCOPIA43* corresponding to cleaved mRNA fractionated by gel electrophoresis. miRNA alignment with 5′ RLM RACE PCR cleavage sites, and frequencies of sequenced cleavage products. Asterisks indicate transcript cleavage site.

(Fig. 2a). *ATHILA ORF1* contains a predicted target site for miR-859, and PARE confirmed cleavage at this site (Fig. 2c and Supplementary Table 3). We further validated miR-859-directed cleavage by modified 5′-RNA-ligase-mediated rapid amplification of cDNA ends polymerase chain reaction (5′ RLM RACE PCR) (Fig. 2e).

Despite cleavage of many TE families by multiple miRNAs (Extended Data Fig. 10), some did not generate easiRNAs. This was especially true of *CACTA* and *ATCOPIA* elements, for which most cleavage events were non-productive. For example, *ATCOPIA43* elements were targeted by miR-390, which targets non-coding RNA for tasiRNA production by the two-hit model[16], but which did not generate easiRNAs from *ATCOPIA43* in *ddm1* (Fig. 2b). Instead, PARE detected uncapped degradation products from *ATCOPIA43*, indicating extensive secondary RNA decay (Fig. 2d) after miRNA cleavage (Fig. 2f), and similar mRNA decay patterns were found at many genes targeted by miRNA (Extended Data Fig. 2). In general, easiRNA-producing TEs were intact *ATENSPM2*, *ATENSPM5*, *ATENSPM6*, *ATCOPIA93* and *ATCOPIA28* elements, whereas those that did not generate easiRNAs were non-autonomous elements (for example, *ATENSPM1A*) interrupted by insertion of other TEs or otherwise truncated, and were subject to RNA decay. Occasionally, TEs were found to produce easiRNAs and were predicted to be targeted by miRNAs, but cleavage could not be detected by PARE, which may be due to the underrepresentation of some miRNAs in inflorescence tissue (Supplementary Table 2). For example, *ATCOPIA93* produces abundant easiRNAs from the *GAG* gene, which is predicted to be targeted by several known miRNAs and eamiRNAs (Extended Data Fig. 3a), yet only miR-833 shows evidence of cleavage by PARE, and this did not pass our cut-off for miRNA cleavage (Supplementary Table 3). *ATCOPIA93* is specifically expressed in only a subset of cells[18].

Two new classes of miRNAs were found in *ddm1*, namely those encoded by transposons and those for which miRNA precursor genes are methylated. We have labelled these miRNAs epigenetically activated (ea)miRNAs, as, like easiRNAs, they are abundant in pollen and *ddm1* (Extended Data Figs 4, 10 and Supplementary Table 2). We also identified new miRNA isomers, changed from 21-nucleotide to 22- and 24-nucleotide sequence variants, originating from known miRNA precursors (Supplementary Table 2). Twenty-two-nucleotide isoforms promote secondary siRNA biogenesis[3,17], whereas 24-nucleotide isoforms promote DNA methylation[9,19]. The newly identified eamiR-2

©2014 Macmillan Publishers Limited. All rights reserved

originates from an immature precursor sequence within an *ATHILA4* retroelement and is abundant in *ddm1* only (Supplementary Table 2). PARE analysis of *ddm1* confirmed release of this eamiRNA from its own *ATHILA4* precursor (Supplementary Table 3), and cleavage of other *ATHILA4* elements *in trans* (Fig. 1). However, this TE-derived eamiRNA does not seem to direct easiRNA biogenesis from its own precursor (Supplementary Table 1). Thus, the release of TE-producing eamiRNAs by DICER does not trigger easiRNA biogenesis *per se*, but only when the miRNA targets TE transcripts via *AGO1* (ref. 7).

To test the requirement for miRNA in easiRNA biogenesis, we used miR-845b, a 22-nucleotide miRNA predicted to target several retroelements (Supplementary Table 3), but only in pollen where it is specifically expressed. We fused a green fluorescent protein (*GFP*) reporter gene to a ubiquitously expressed promoter, and to a 300 base pair (bp) region of *ATGP1* that included the predicted miR-845b target site (Fig. 3a). Sequencing small RNA from pollen revealed novel 21-nucleotide easiRNAs surrounding the miR-845b target site (Fig. 3a), which were not found in pollen with constructs in which the miRNA target site was deleted (Fig. 3b). We confirmed by modified 5′ RLM RACE PCR that miR-845b-directed cleavage products from *GFP-ATGP1* transcripts accumulate exclusively in pollen (Fig. 3c). Thus, 21-nucleotide easiRNA biogenesis at TEs depends on targeting by miRNA.

Twenty-four-nucleotide hetsiRNAs guide asymmetric CHH methylation at TEs and are *RDR2* dependent[4]. We profiled 24-nucleotide hetsiRNAs in Col-0, *ddm1*, *rdr6*, *ddm1 rdr6* and *ddm1 dcl1-11* (Supplementary Table 1 and Fig. 4). In Col-0, 24-nucleotide hetsiRNAs target LTR retrotransposons and most DNA transposons (Extended Data Fig. 5). In *ddm1*, we found a slight increase in multiple mapping 24-nucleotide hetsiRNAs from *ATGYPSY* retrotransposons that were saturated by easiRNA biogenesis (Supplementary Table 1), but an overall general loss from DNA transposons and other individual retrotransposons, such that less than half of these retained 24-nucleotide siRNA in *ddm1* (Extended Data Fig. 5). Many of these TEs gained 24-nucleotide hetsiRNAs in *ddm1 rdr6* and *ddm1 dcl1* (Supplementary Table 1 and Extended Data Fig. 5) and, furthermore, hundreds of TEs that did not have detectable hetsiRNAs in Col-0 or *ddm1* gained them in *ddm1 rdr6* (Fig. 4a and Extended Data Fig. 6). Thus, miRNA targeting and easiRNA biogenesis might inhibit 24-nucleotide hetsiRNA production from transposons in *ddm1*. We found that almost all TEs cleaved by miRNA either gained 21-nucleotide easiRNAs in *ddm1*, lost 24-nucleotide hetsiRNAs in *ddm1*, or gained 24-nucleotide hetsiRNAs in *ddm1 rdr6* (Fig. 4b, c and Extended Data Fig. 6). This enrichment

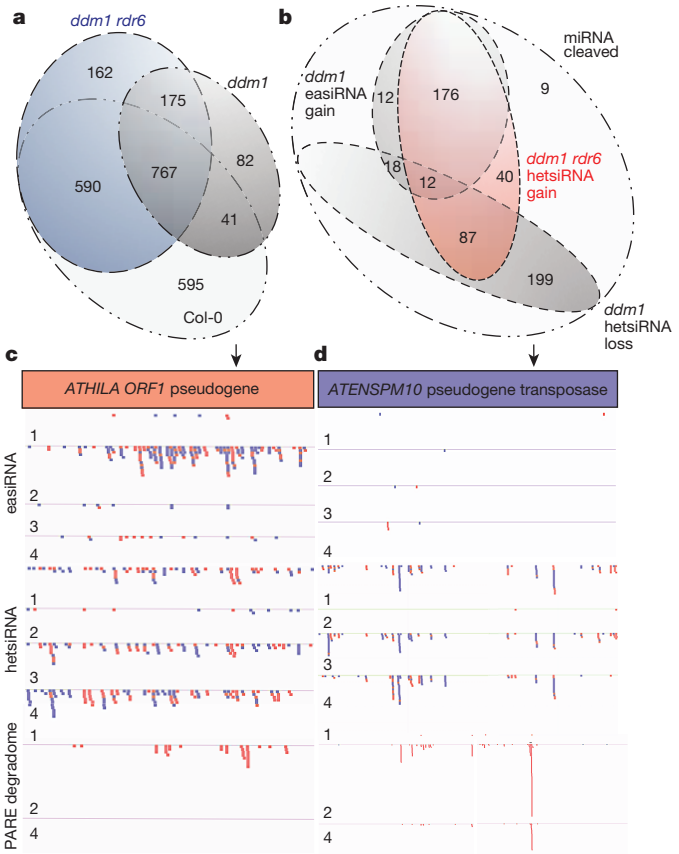


Figure 4 | Transposons targeted by miRNA gain hetsiRNA when the easiRNA pathway is lost. **a**, Overlap of transposons targeted by hetsiRNAs in Col-0, *ddm1-2* and *ddm1-2 rdr6-15*. **b**, Overlap of transposons targeted by miRNAs that have lost hetsiRNAs in *ddm1-2*, gained hetsiRNAs in *ddm1-2 rdr6-15*, and those whose transcripts undergo productive miRNA cleavage resulting in easiRNA biogenesis in *ddm1-2*. **c**, **d**, Twenty-one-nucleotide easiRNAs, 24-nucleotide hetsiRNAs and PARE degradome reads from easiRNA-generating *ATHILA* (AT3G32118) (**c**) in comparison with non-easiRNA generating *ATENSPM1* (AT4G02314) (**d**). Individual reads: sense, red; antisense, blue. Track order: 1, Col-0; 2, *ddm1-2*; 3, *rdr6-15*; 4, *ddm1-2 rdr6-15*.

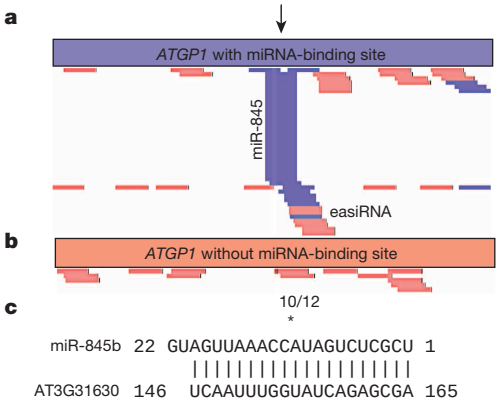


Figure 3 | miR-845b targets *ATGP1*, promoting easiRNA biogenesis. **a**, Transgenic line overexpressing *ATGP1* containing the 20-nucleotide miR-845b target region (marked miR-845) and novel overlapping 21-nucleotide easiRNAs (marked easiRNA). Individual reads: sense, red; antisense, blue. **b**, Transgenic line overexpressing *ATGP1* that did not contain the 20-nucleotide miR-845b predicted target region, with no 21-nucleotide easiRNAs overlapping this site. **c**, miRNA alignment with 5′ RLM RACE PCR cleavage sites, and frequencies of sequenced cleavage products. Asterisk indicates miR-845b:*ATGP1* cleavage site.

was even greater for those TEs targeted by more than one miRNA (Extended Data Fig. 6c).

Thus, miRNA cleavage seems to inhibit 24-nucleotide siRNA biogenesis in a manner that depends on RDR6, whether or not easiRNAs accumulate (Fig. 4b and Supplementary Table 3). This response was observed when miRNA cleavage sites were on the sense (mRNA) or occasionally on the antisense strand (Supplementary Table 3). It is possible, therefore, that miRNA-directed cleavage of antisense Pol IV, or Pol V transcripts, and processing by *RDR6* rather than *RDR2*, might prevent the formation of 24-nucleotide siRNAs without necessarily generating 21-nucleotide easiRNAs (that are Pol II dependent[8]). Thus, *RDR6* partially antagonizes *RDR2* activity, inhibiting 24-nucleotide hetsiRNA biogenesis at TEs. Likewise, *RDR2* partially antagonizes *RDR6* activity at transgenes subject to PTGS[20].

Symmetric CG and CHG methylation contexts are maintained by DNA methyltransferases and histone modifications, whereas CHH methylation is associated with 24-nucleotide hetsiRNA-guided RNA-directed DNA methylation (RdDM)[4]. Methylation mediates transcriptional silencing of TE promoters[21], found near the terminal inverted repeats (TIR) of most DNA transposons and in the LTR of retrotransposons, as well as internally in the case of the *ATHILA ORF1* and some DNA transposons. As expected, whole-genome sequencing of bisulphite-treated DNA from inflorescence tissue revealed global loss of DNA

©2014 Macmillan Publishers Limited. All rights reserved

methylation from all classes of transposons in *ddm1* when compared to wild type (Extended Data Fig. 7a, b). As previously reported, most of this loss occurred in the symmetric CG and CHG contexts, rather than the asymmetric CHH context, consistent with retention of 24-nucleotide hetsiRNAs[22] (Supplementary Table 4). By using whole-genome bisulphite sequencing, we measured levels of DNA methylation in *ddm1 rdr6* F3 progeny from two independent double mutants relative to *ddm1* (Extended Data Fig. 7a). A modest difference in DNA methylation is expected due to inbreeding of the *ddm1* parents[23], and was observed in one *ddm1 rdr6* replicate; however, methylation at TEs substantially increased in the other replicate (Extended Data Fig. 7e, f and Supplementary Table 4). These results suggest that elevated levels of 24-nucleotide hetsiRNAs could only guide remethylation of TEs stochastically in *ddm1 rdr6* (Extended Data Fig. 7g, h), probably reflecting the epigenetic inheritance of unmethylated and methylated TEs from the *ddm1* and *rdr6* parents, respectively, as previously observed[24].

To assess whether the loss of the easiRNA pathway leads to a further outburst of transposon reactivation in *ddm1 rdr6*, we used Affymetrix *ATH1* microarrays to analyse the transcriptome. A few hundred representative TEs are included on Affymetrix *ATH1* microarrays, and can be used to assess TE reactivation[5]. We found that most of these TE transcripts were abundantly and similarly expressed in both *ddm1* and *ddm1 rdr6* (Supplementary Table 5 and Extended Data Fig. 8), indicating that easiRNA biogenesis does not reduce TE transcript levels in *ddm1* mutants. In fact, consistent with our findings of sporadic TE methylation and gain of hetsiRNAs in *ddm1 rdr6,* several TEs had reduced expression in *ddm1 rdr6*, including *ATHILA* and *ATCOPIA* elements, as well as TEs of the MuDR superfamily (Supplementary Table 5). Therefore, the loss of both *DDM1* and *RDR6* does not enhance transcriptional reactivation of TEs, but instead, some TEs become transcriptionally repressed in *ddm1 rdr6* (Extended Data Fig. 8 and Supplementary Table 5) and an elevation in methylation level is observed at least for some of these TEs, for example, *ATCOPIA22* and *ARNOLD3*, in *ddm1 rdr6* (Supplementary Table 4).

We have found that loss of TE methylation results in epigenetic activation and consequent transcription, whereby transposon mRNA become preferentially targeted by more than 50 miRNAs bound to *AGO1*. Productive cleavage of transposon transcripts usually engages *RDR6*, allowing *DCL4* to generate 21-nucleotide easiRNAs in a PTGS mechanism (Extended Data Fig. 9). About half of these transposons belong to the LTR/*ATHILA* element family (Extended Data Fig. 10), which occur in high copy number at pericentromeric repeats and are targeted by multiple miRNAs (Extended Data Fig. 3c and Fig. 2). Several known miRNA genes—for example, *MIR843*—are methylated, and only target transposon transcripts when unmethylated and expressed: in pollen and in *ddm1* mutants (Supplementary Table 2 and Extended Data Fig. 4). Other miRNAs are developmentally regulated—for example, miR-156, which is required to maintain the juvenile phase of plant development, and miR-172, which is required for the transition to the adult phase[25]—accounting perhaps for the tissue-specific silencing of TEs during different phases of plant development[26].

TEs are targeted by the same conserved miRNA (miR-845) in rice and *Arabidopsis*, leading to the idea that they may have retained miRNA-binding sites despite selection against them. Furthermore, some of the newly identified miRNAs are themselves encoded by TEs (Extended Data Fig. 10). It is unclear why transposons would retain miRNA sites that could potentially silence them post-transcriptionally. One explanation lies in the preferential processing of targeted TE transcripts via *RDR6*, which seems to act antagonistically with *RDR2*, preventing 24-nucleotide hetsiRNA biogenesis[20] (Extended Data Fig. 9). We found a notable overlap between TEs that produced easiRNAs in *ddm1* and those that gained 24-nucleotide hetsiRNAs in *ddm1 rdr6* (Fig. 4b and Extended Data Fig. 6). Twenty-four-nucleotide hetsiRNAs promote RdDM, and it is likely, therefore, that TEs tolerate developmentally regulated PTGS by miRNA and easiRNAs to avoid long-term TGS by DNA methylation. Given the probable evolution of miRNA precursors from TEs and other inverted repeats[21], miRNAs may have arisen in ancient eukaryotes to target retrotransposons and other TEs rather than genes.

ddm1 mutants are remarkably normal, despite the heritable loss of heterochromatin. By contrast, *ddm1 rdr6*, *ddm1 dcl4* and *ddm1 dcl1* have a wide range of developmental phenotypes and heritable epigenetic defects, including widespread infertility in subsequent generations (K.M.C. and R.A.M., unpublished observations). We therefore postulate that the function of the easiRNA biogenesis pathway within the host is to protect the genome from TE-mediated epigenomic instability. Consistent with this idea, easiRNAs specifically target the *ATHILA ORF1* gene, thought to be involved in retrovirus-like particle formation, as well as the *ATCOPIA* integrase gene and the *CACTA* transposase gene, required for transposition (Extended Data Fig. 3b and Fig. 4c). *RDR6*-directed PTGS could be dangerous to the host by presenting an opportunity to the TE for evasion and transposition, whereas *RDR2*-directed TGS is a safer silencing strategy not requiring transcription. An important developmental stage during which PTGS is deployed is in the pollen grain, when pericentromeric retrotransposons, and some DNA transposons, lose CHH methylation and RdDM in the male germ line[27]. Many of these same TEs accumulate easiRNAs in sperm, which are generated by TE activation in the companion vegetative nucleus, accompanied by the loss of DDM1 (ref. 5). Transient loss of RdDM in sperm is restored after fertilization by maternal 24-nucleotide hetsiRNAs[27,28] allowing an opportunity to distinguish 'self' from 'non-self' pollen on fertilization, according to whether the transposon load is sufficiently foreign not to be recognized by maternal small RNAs. PTGS, mediated by easiRNAs, would provide a backup silencing mechanism during this critical but potentially dangerous window. A similar secondary siRNA transposon control mechanism, triggered by PIWI-interacting RNA (piRNA) rather than miRNA, has recently been described in *Caenorhabditis elegans* and may have similar roles[29,30].

METHODS SUMMARY

Genomic DNA and total RNA were isolated from inflorescence tissue (stage 9–11) collected from Col-0 wild type and loss of function mutants: *ddm1-2*, *rdr6-15*, *ddm1-2 rdr6-15* and *ddm1-2 dcl1-11* in a Col-0 background, grown under standard long-day growth conditions. Small RNA libraries were sequenced on Illumina Hiseq and Miseq and mapped using Bowtie from the open-source Tuxedo suite, allowing for both unique (U) and multiple (M) mapping reads. Bisulphite sequencing (BIS-seq) library construction and bisulphite conversion were performed as previously described. Col-0, *ddm1-2*, *rdr6-15* and *ddm1-2 rdr6-15* replicate BIS-seq libraries were sequenced on Illumina HiSeq2000 using paired-end reads for 50 cycles, and mapped using BISmark. Our miRNA prediction algorithm was used to identify miRNAs that target TEs. mRNA purification (Dynabeads), 5′ adaptor ligation (Invitrogen) and cDNA purification (AMPure XP) for PARE were performed according to the manufacturer's instructions, followed by PCR amplification and PAGE purification. 5′ RLM RACE was performed according to the manufacturer's instructions (Ambion). Sequencing and mapping information for each library described in this study is given in Supplementary Table 8.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 January 2013; accepted 22 January 2014.
Published online 16 March 2014.

1. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121,** 207–221 (2005).
2. Ronemus, M., Vaughn, M. W. & Martienssen, R. A. MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent RNA polymerase in *Arabidopsis. Plant Cell* **18,** 1559–1574 (2006).
3. Cuperus, J. T. *et al.* Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis. Nature Struct. Mol. Biol.* **17,** 997–1003 (2010).
4. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11,** 204–220 (2010).
5. Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136,** 461–472 (2009).

©2014 Macmillan Publishers Limited. All rights reserved

6. Tanurdzic, M. *et al.* Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol.* **6,** e302 (2008).
7. McCue, A. D., Nuthikattu, S., Reeder, S. H. & Slotkin, R. K. Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet.* **8,** e1002474 (2012).
8. Nuthikattu, S. *et al.* The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.* **162,** 116–131 (2013).
9. Mirouze, M. *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461,** 427–430 (2009).
10. Miura, A. *et al.* Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411,** 212–214 (2001).
11. German, M. A., Luo, S., Schroth, G., Meyers, B. C. & Green, P. J. Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nature Protocols* **4,** 356–362 (2009).
12. Hsieh, T. F. *et al.* Genome-wide demethylation of *Arabidopsis* endosperm. *Science* **324,** 1451–1454 (2009).
13. Gehring, M. *et al.* DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124,** 495–506 (2006).
14. Schwab, R., Speth, C., Laubinger, S. & Voinnet, O. Enhanced microRNA accumulation through stemloop-adjacent introns. *EMBO Rep.* **14,** 615–621 (2013).
15. Jeong, D. H. *et al.* Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* **23,** 4185–4207 (2011).
16. Axtell, M. J. & Bartel, D. P. Antiquity of microRNAs and their targets in land plants. *Plant Cell* **17,** 1658–1673 (2005).
17. Chen, H. M. *et al.* 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl Acad. Sci. USA* **107,** 15269–15274 (2010).
18. Marí-Ordóñez, A. *et al.* Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genet.* **45,** 1029–1039 (2013).
19. Wu, L. *et al.* DNA methylation mediated by a microRNA pathway. *Mol. Cell* **38,** 465–475 (2010).
20. Jauvion, V., Rivard, M., Bouteiller, N., Elmayan, T. & Vaucheret, H. RDR2 partially antagonizes the production of RDR6-dependent siRNA in sense transgene-mediated PTGS. *PLoS ONE* **7,** e29785 (2012).
21. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature Genet.* **37,** 641–644 (2005).
22. Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152,** 352–364 (2013).
23. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22,** 94–97 (1999).
24. Teixeira, F. K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323,** 1600–1604 (2009).
25. Poethig, R. S. Small RNAs and developmental timing in plants. *Curr. Opin. Genet. Dev.* **19,** 374–378 (2009).
26. Poethig, R. S. Phase change and the regulation of shoot morphogenesis in plants. *Science* **250,** 923–930 (1990).
27. Calarco, J. P. *et al.* Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151,** 194–205 (2012).
28. Mosher, R. A. *et al.* Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* **460,** 283–286 (2009).
29. Bagijn, M. P. *et al.* Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* **337,** 574–578 (2012).
30. Shirayama, M. *et al.* piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* **150,** 65–77 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank V. Colot, O. Voinnet and A. Sarazin for sharing unpublished data and for discussions, J. Simorowski for plant genetics and J. Kendall for computational advice. We thank R.K. Slotkin for sharing data before publication. K.M.C. and M.R. were supported, in part, by a research collaboration with DuPont Pioneer. F.V.E. was supported by a fellowship from the Belgian American Educational Foundation. J.Z. was supported by a University of Delaware Graduate fellowship. Research in the Martienssen laboratory is supported by a grant from the National Institutes of Health (RO1GM067014 to R.A.M.) and by the Howard Hughes Medical Institute and Gordon and Betty Moore Foundation (GBMF3033). The authors acknowledge that this work was performed with assistance from the Cold Spring Harbor Laboratory Shared Resources, which are funded, in part, by the Cancer Center Support Grant (5PP30CA045508).

Author Contributions R.A.M., B.C.M. and K.M.C. designed the study. F.B., F.V.E., M.R. and K.M.C. performed experiments. J.Z. and K.M.C. analysed data. R.A.M. and K.M.C. prepared the manuscript.

Author Information Data have been deposited in the Gene Expression Omnibus under accession numbers GSE52952, GSE52067, GSE52342, GSE52346 and GSE52951. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.M. (martiens@cshl.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13037

Structural basis for translocation by AddAB helicase–nuclease and its arrest at χ sites

Wojciech W. Krajewski[1]†, Xin Fu[1], Martin Wilkinson[1], Nora B. Cronin[1], Mark S. Dillingham[2] & Dale B. Wigley[1]

In bacterial cells, processing of double-stranded DNA breaks for repair by homologous recombination is dependent upon the recombination hotspot sequence χ (Chi)[1,2] and is catalysed by either an AddAB- or RecBCD-type helicase–nuclease (reviewed in refs 3, 4). These enzyme complexes unwind and digest the DNA duplex from the broken end until they encounter a χ sequence[5], whereupon they produce a 3′ single-stranded DNA tail onto which they initiate loading of the RecA protein[6]. Consequently, regulation of the AddAB/RecBCD complex by χ is a key control point in DNA repair and other processes involving genetic recombination. Here we report crystal structures of *Bacillus subtilis* AddAB in complex with different χ-containing DNA substrates either with or without a non-hydrolysable ATP analogue. Comparison of these structures suggests a mechanism for DNA translocation and unwinding, suggests how the enzyme binds specifically to χ sequences, and explains how χ recognition leads to the arrest of AddAB (and RecBCD) translocation that is observed in single-molecule experiments[7–9].

We have previously determined the crystal structures of initiation complexes of AddAB and RecBCD bound to a DNA end[10,11]. We have now crystallized AddAB in a translocation-like state by extending the 5′ and 3′ tails of the substrate to mimic an unwound fork (Fig. 1 and Extended Data Fig. 1). These structures provide additional details of the interaction between the AddA helicase domain and the 3′ tail of the DNA fork (Extended Data Fig. 2). The structure of the complex with the non-hydrolysable ATP analogue ADPNP revealed that AddAB contains two ATP-binding sites (Fig. 1a). One is located between the helicase domains of AddA, and the other is in the equivalent position between the catalytically inactivated helicase domains of AddB. The ATP-binding site in AddA is required for processive DNA translocation and unwinding[12], and is essentially the same as that observed in other SF1 helicases[13]. The ATP-binding site in AddB is important for stabilizing the complex between AddAB and χ[12]. It contains several residues that are similar to those in AddA, and ADPNP is bound in a similar fashion at both sites (Fig. 1b and Extended Data Fig. 3). However, there are important differences in residues associated with hydrolysis rather than binding of ATP. For example, essential catalytic glutamate and arginine residues (E408 and R873 in AddA) are replaced in AddB by a glycine and a serine, respectively. This suggests that AddB has a much reduced ATPase activity compared to AddA, consistent with biochemical studies[12]. Conservation of an ATP-binding site in AddB further substantiates proposals that the χ-scanning subunits of RecBCD and AddAB (RecC and AddB, respectively) have evolved from bona fide helicase subunits[4,14].

The binding of nucleotide to the AddA subunit induces conformational changes in the motor domains (1A and 2A) comparable to those seen for other SF1 and SF2 helicases[13] (Fig. 1a). By contrast, the conformation of the AddB subunit remains unaltered by the binding of ADPNP other than local changes in amino-acid side chains. The conformational changes in AddA suggest both a mechanism for single-stranded DNA (ssDNA) translocation, similar to that proposed for other SF1 helicases[13,15], and how the complex might unwind DNA (Fig. 2). Immediately ahead of the ssDNA motor, the 'arm' domain of AddA and the carboxy-terminal nuclease domain of AddB contact the duplex on opposite sides such that they almost encircle the DNA. Upon ATP binding, the 1A and 2A motor domains of AddA move together, causing ssDNA to slide by one base across the surface of the 1A domain and the 3′ ssDNA tail to be pulled away from the DNA junction. At the same time, the arm domain pulls the duplex in the opposite direction, thereby creating tension that is relieved by unwinding a base pair at the fork junction. When ATP is hydrolysed, the motor domains revert to the open state, domain 1A slides back along the ssDNA towards the junction and the regions of the protein contacting the DNA duplex also slide backwards, to alter the register of DNA contacts by one base pair (see Supplementary Videos 1 and 2). Hence one

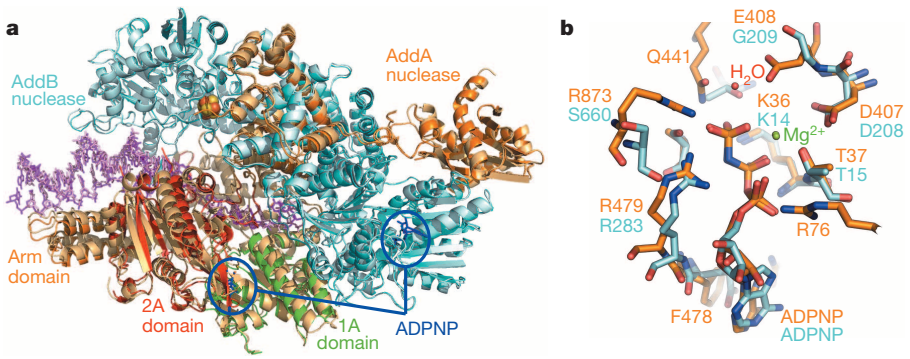


Figure 1 | Structure of the binary and ternary complexes. a, Comparison of the binary and ternary (ADPNP) complexes. AddA is shown in beige (binary) and orange (ADPNP) but with the motor domains of AddA in the ADPNP complex in green and red, AddB in light cyan (binary) and cyan (ADPNP), the DNA is shown in pink (binary) and purple (ADPNP), bound ADPNP molecules in blue sticks, and the iron–sulphur cluster as spheres. **b**, Overlay of ATP-binding sites in AddA (orange) and AddB (cyan). Although a bound magnesium ion is superimposable in both sites, a bound water molecule, positioned for attack of the γ-phosphate, is only evident in the AddA site.

[1]Division of Structural Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK. [2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK. †Present address: CRT Discovery Laboratories, Department of Biological Sciences, Birkbeck, University of London, London WC1E 7HX, UK.

©2014 Macmillan Publishers Limited. All rights reserved

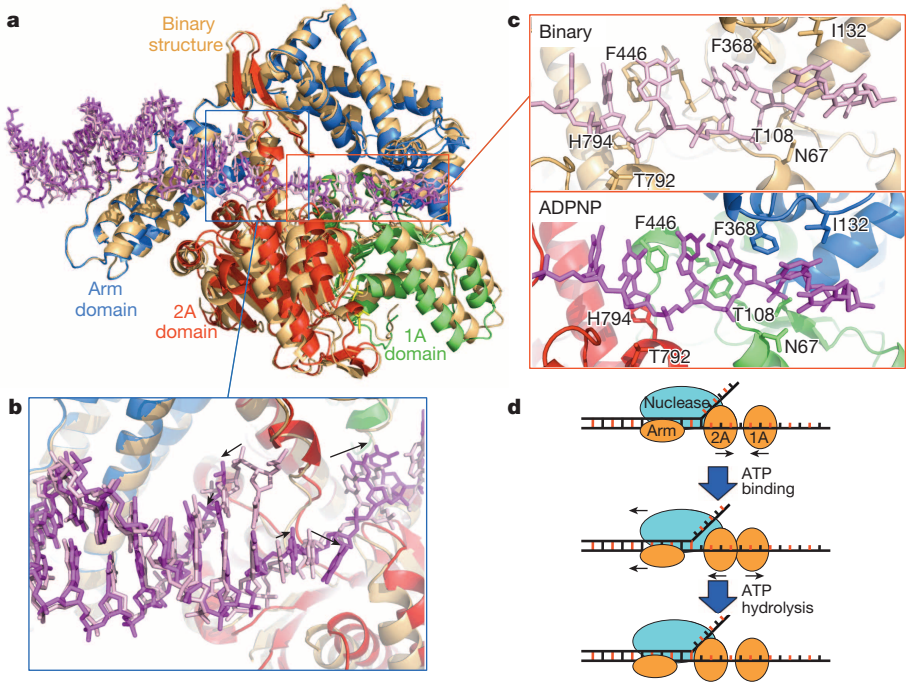


Figure 2 | Proposed mechanism of unwinding and translocation. **a**, Superposition of the structures on the arm domain illustrates the stretching of the DNA junction that occurs when ATP-binding induces movement of the 1A and 2A domains. For clarity, only the DNA and domains 1A, 1B and 2A are shown. The binary complex is coloured in beige (protein) and pink (DNA), whereas the ADPNP structure is coloured as labelled in the figure, with the DNA in purple. **b**, The register of contacts between (1) the arm domain and DNA duplex; and (2) the 2A domain and ssDNA remains the same but their separation increases upon ATP binding. The strain in the duplex is relieved by unwinding of the duplex causing disruption of a single base pair at the junction. Arrows highlight some of the important movements of the DNA. **c**, Contacts with the ssDNA tail change register by a single base upon ATP binding as the DNA slides across the surface of domain 1A. **d**, Cartoon summarizing the minimal steps in the unwinding mechanism. Upon ATP binding, the 1A and 2A domains of AddA close and the conformational change slides along the 3′ ssDNA tail across the surface of the 1A domain by one base. At the same time, the spacing between the arm domain and the 2A domain increases, introducing stress into the duplex and causing it to unwind a base pair at the junction. After hydrolysis of ATP, the 1A and 2A domains relax to the open conformation, causing the 2A domain to slide across the ssDNA. At the same time, the arm domain and the nuclease domain of AddB slide by one base pair along the duplex. This simple two-step process identifies a minimal set of conformational changes that are required to unwind the DNA but does not preclude other intermediates.

cycle of ATP binding and hydrolysis unwinds and translocates the complex by one base pair in common with many other helicases[13]. The unexpected and direct role for the arm domain in this mechanism may provide a molecular basis for the enigmatic secondary translocase activity detected in the structurally related RecBC complex[16].

The AddB subunit is responsible for χ recognition[10] but the molecular details of this interaction are unknown. To mimic delivery of χ to the recognition locus, we next prepared a series of DNA substrates that placed a *B. subtilis* χ sequence (5′-AGCGG (ref. 2)) on the 3′ strand at increasing distances from the junction by gradually increasing the length of a 'spacer' region (Extended Data Fig. 1). These substrates all crystallized with AddAB and ADPNP in the same conditions, space group and unit cell as those lacking χ. The overall conformation of the protein was the same in all structures, with no electron density attributable to a bound χ site, with one exception. In the case of the substrate with a seven-base spacer, there was strong electron density within the AddB subunit corresponding to a bound χ sequence (Fig. 3) as well as substantial alterations in the conformation of the AddA motor domains (discussed below).

Both RecC and AddB have the same fold as an SF1 helicase[10,11] and in AddB even the ATP-binding site is conserved, albeit with changes that probably reduce or abolish ATPase activity (see above). Furthermore, some residues known to affect binding and/or response to χ in AddAB and RecBCD are located in positions equivalent to the characteristic helicase motifs[10,17,18]. Consequently, we expected the DNA to bind in a manner similar to that observed for SF1 DNA helicases[13]. The χ-binding site is indeed located in the same groove as the ssDNA-binding site in SF1 helicases (Fig. 3a, b). Moreover, residues either side of the χ sequence interact with the protein principally via the phosphodiester backbone, which is largely similar to ssDNA binding in other DNA helicases that need to bind DNA in a non-sequence-specific manner[13]. However, the manner in which the five bases of the χ sequence interact with the protein is entirely different. Starting with the first residue (adenine), the DNA is flipped by 180° such that the bases point towards, rather than away from, the protein surface. As a result the entire interface with χ is via the bases, but bounded by the phosphate groups on either side that also make important contacts (see below). This mode of binding is thereby optimized for specificity over affinity: any interaction with the backbone would be sequence-independent, increasing affinity for all DNA sequences rather than specificity for χ. All five bases of the χ sequence interact with the protein (Fig. 3c). Several residues in AddB that either contact the χ site directly (Q42, T44, R70, F210) or order residues that contact the bases (F68, W73) had been implicated previously in the recognition or response to χ (ref. 10). However, this structure reveals a role for several additional residues, including some, surprisingly, that are in the AddA subunit (Fig. 3c). Many of these have equivalents in RecB and RecC[17], probably reflecting a commonality in the mechanism of χ regulation in AddAB and RecBCD.

The χ-binding site extends across from the AddB subunit into the AddA nuclease domain, which has the same fold as λ-exonuclease[19]. The structure of λ-exonuclease complexed with DNA[20] identified the importance of the 5′ phosphate of the substrate for both binding and catalysis. The function of an arginine residue (R28) that has a key role in this contact is replaced by a tyrosine (Y1204) in AddA (Fig. 3d). However, in λ-exonuclease, additional contacts with this phosphate moiety involve residues T33–S35, a region that is conserved in the nuclease domain of AddA (S1015–S1017) within the linker connecting the helicase and nuclease domains. These residues coordinate the equivalent

©2014 Macmillan Publishers Limited. All rights reserved

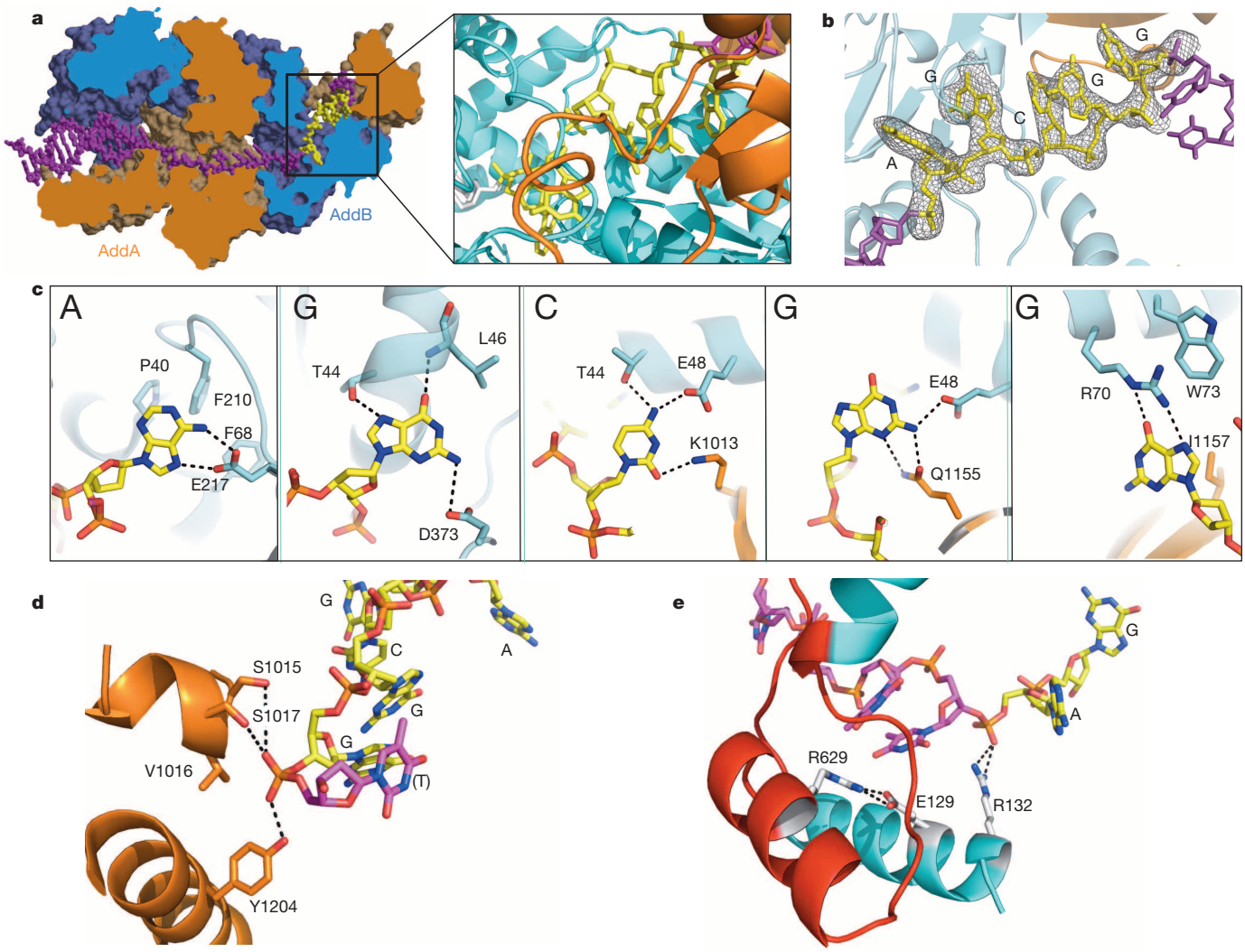


Figure 3 | The complex of AddAB bound to χ. **a**, Cutaway space-filling representation of the ADPNP–χ complex showing the long channel through the complex, along which the DNA runs. DNA is purple but with the χ residues in yellow (inset shows close up of the χ site but in ribbon format). **b**, Difference electron density ($F_o - F_c$, contoured at 2.5σ) corresponding to the five residues of the χ site, with the final structure overlaid for reference. **c**, Details of interactions with each base of the χ sequence. **d**, Binding site for the phosphate group at the 3′ end of the χ sequence (carbon atoms yellow). **e**, Link between the phosphate binding site at the 5′ end of the χ residues (yellow) and the latch residues (red).

position in the phosphodiester backbone on the 3′ side of the final G of the χ sequence. A similar motif is conserved in the AddB nuclease domain and probably contributes in an equivalent manner to binding of the 5′ tail. As the 3′ ssDNA tail continues beyond the χ sequence and enters the AddA nuclease domain, the conformation flips back by 180° so that the principal contacts are once again with the phosphodiester backbone rather than the bases.

Single-molecule experiments with RecBCD[7,8] and AddAB[9] revealed that both complexes show interesting behaviour when encountering a χ sequence. In each case, the complexes are unwinding DNA at several hundred base pairs per second before suddenly pausing at χ for several seconds. After this pause, the complexes resume translocation but usually at a lower speed and with modified properties such that the complex is no longer able to recognize a second χ site, suggesting that the enzyme remains associated with the original χ site, thereby blocking binding in a subsequent encounter. Initially, two explanations for the pause were proposed for RecBCD[7]. The first was that, because the RecB and RecD motors run at different speeds[21] and the RecD motor becomes inactivated at χ[22] (although it remains associated with the complex[23,24]), the pause at χ was a consequence of the slower (RecB) motor 'catching up' with additional bases unwound by the faster RecD before net translocation by RecB could recommence. The change in translocation speed observed for RecBCD was attributed to a change in the nature of the lead motor from the faster RecD before χ to the slower RecB after χ. However, subsequent work showed that a mutant RecBCD enzyme with an inactivated RecD subunit displays the same pausing behaviour[8], as does the single-motor AddAB complex[9], ruling this out as an explanation.

The second proposal was that a χ-dependent conformational change takes place during the pause that is required before translocation can recommence[7,23]. During the pause, the enzyme is stalled in an altered, but unknown, conformational state until it reactivates to a form that can recommence translocation. However, no information about this χ-dependent stalled conformation could be obtained from the single-molecule experiments.

Our structure of the χ-bound state of AddAB now reveals the nature of the conformational change that takes place upon encountering χ. Unexpectedly, the 1A and 2A motor domains of the AddA subunit adopt the conformation for the complex without bound nucleotide, despite being grown under the same conditions as the ADPNP complex and crystallizing in the same space group and unit cell. Importantly, the ATP-binding sites of AddA and AddB are both occupied with a bound ADPNP (Extended Data Fig. 3). Consequently, when χ is bound to the AddB subunit, the AddA motor domains are unable to adopt the usual nucleotide-bound closed conformation, thereby blocking

©2014 Macmillan Publishers Limited. All rights reserved

DNA translocation and explaining the pause seen in single-molecule experiments[7–9]. We have now determined several structures of χ-bound complexes from different crystals and these show that although domain 1A is consistently in an orientation similar to that in the binary complex, the 2A domain adopts a range of non-canonical conformations (Extended Data Fig. 4), suggesting that domain 1A is locked but domain 2A becomes uncoupled and is able to adopt multiple states. Hence, the paused χ-bound state may well remain able to hydrolyse ATP without translocating, rather like a motor idling without the gearbox engaged, until a subsequent conformational change allows the complex to re-engage and recommence translocation.

Although our structure reveals how the pause is initiated, the reason for the pause remains a point for speculation. One possibility is that the protein requires a slow conformational change to take place before the complex is proficient for the next phase of the reaction, namely the loading of RecA protein. Data for both RecBCD and AddAB suggest this may involve the opening of a protein gateway to allow extrusion of an ssDNA loop[10,17,25]. Interestingly, the structure reveals a direct link from the χ-binding site (R132) to one of the residues (E129) that forms part of a proposed latch (Fig. 3e) responsible for the opening of this exit channel, although the latch remains in the closed conformation in this structure. It might also be that the final DNA cleavage event on the χ-containing strand has to take place before translocation resumes and RecA begins to load. This would act as a failsafe mechanism to ensure the production of an appropriate substrate for extension by the action of DNA polymerase following RecA-dependent strand invasion. Consistent with this view, a nuclease-defective RecBCD complex is not able to load RecA protein at χ[26]. The work we present here captures several intermediates on the AddAB repair pathway (Extended Data Fig. 5), but further work is required to understand how the complex is reactivated following the stall at χ, thereby resuming DNA unwinding and the loading of RecA protein to initiate homologous recombination.

METHODS SUMMARY

Protein and DNA substrates were prepared as described previously[10,11]. Protein–DNA complex was prepared by mixing a slight excess of DNA with protein complex and then purified by gel filtration before crystallization by vapour diffusion. Crystals were flash-frozen in liquid nitrogen before data collection. The structures were determined by molecular replacement using the AddAB coordinates (PDB accession number 3U4Q)[10]. Details of the final models are presented in Extended Data Table 1.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 October 2013; accepted 16 January 2014.
Published online 16 March 2014.

1. Lam, S. T., Stahl, M. M., McMilin, K. D. & Stahl, F. W. Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* **77,** 425–433 (1974).
2. Chédin, F., Noirot, P., Biaudet, V. & Ehrlich, S. D. A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*. *Mol. Microbiol.* **29,** 1369–1377 (1998).
3. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72,** 642–671 (2008).
4. Wigley, D. B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nature Rev. Microbiol.* **11,** 9–13 (2013).
5. Ponticelli, A. S., Schultz, D. W., Taylor, A. F. & Smith, G. R. Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell* **41,** 145–151 (1985).
6. Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90,** 77–86 (1997).
7. Spies, M. *et al.* A molecular throttle: the recombination hotspot chi controls DNA translocation by the RecBCD helicase. *Cell* **114,** 647–654 (2003).
8. Spies, M., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. RecBCD enzyme switches lead motor subunits in response to chi recognition. *Cell* **131,** 694–705 (2007).
9. Carrasco, C., Gilhooly, N. S., Dillingham, M. S. & Moreno-Herrero, F. On the mechanism of recombination hotspot scanning during double-stranded DNA break resection. *Proc. Natl Acad. Sci. USA* **110,** E2562–E2571 (2013).
10. Saikrishnan, K. *et al.* Insights into Chi recognition from the structure of an AddAB-type helicase-nuclease complex. *EMBO J.* **31,** 1568–1578 (2012).
11. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432,** 187–193 (2004).
12. Yeeles, J. T., Gwynn, E. J., Webb, M. R. & Dillingham, M. S. The AddAB helicase-nuclease catalyses rapid and processive DNA unwinding using a single Superfamily 1A motor domain. *Nucleic Acids Res.* **39,** 2271–2285 (2011).
13. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76,** 23–50 (2007).
14. Unciuleac, M. C. & Shuman, S. Characterization of the mycobacterial AdnAB DNA motor provides insights into the evolution of bacterial motor-nuclease machines. *J. Biol. Chem.* **285,** 2632–2641 (2010).
15. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97,** 75–84 (1999).
16. Wu, C. G., Bradford, C. & Lohman, T. M. *Escherichia coli* RecBC helicase has two translocase activities controlled by a single ATPase motor. *Nature Struct. Mol. Biol.* **17,** 1210–1217 (2010).
17. Handa, N. *et al.* Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme. *Proc. Natl Acad. Sci. USA* **109,** 8901–8906 (2012).
18. Yang, L. *et al.* Mutation of χ-recognition by RecBCD reveals a regulated molecular latch and suggests a channel-bypass mechanism for biological control. *Proc. Natl Acad. Sci. USA* **109,** 8907–8912 (2012).
19. Kovall, R. & Matthews, B. W. Toroidal structure of λ-exonuclease. *Science* **277,** 1824–1827 (1997).
20. Zhang, J., McCabe, K. A. & Bell, C. E. Crystal structures of λ exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc. Natl Acad. Sci. USA* **108,** 11872–11877 (2011).
21. Taylor, A. F. & Smith, G. R. RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423,** 889–893 (2003).
22. Thaler, D. S. *et al.* Recombination of bacteriophage λ in *recD* mutants of *Escherichia coli*. *Genome* **31,** 53–67 (1989).
23. Dohoney, K. M. & Gelles, J. χ-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature* **409,** 370–374 (2001).
24. Handa, N., Bianco, P. R., Baskin, R. J. & Kowalczykowski, S. C. Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after χ recognition. *Mol. Cell* **17,** 745–750 (2005).
25. Wong, C. J., Rice, R. L., Baker, N. A., Ju, T. & Lohman, T. M. Probing 3′-ssDNA loop formation in *E. coli* RecBCD/RecBC-DNA complexes using non-natural DNA: a model for ''Chi'' recognition complexes. *J. Mol. Biol.* **362,** 26–43 (2006).
26. Anderson, D. G., Churchill, J. J. & Kowalczykowski, S. C. A single mutation, $RecB_{D1080A}$ eliminates RecA protein loading but not Chi recognition by RecBCD enzyme. *J. Biol. Chem.* **274,** 27139–27144 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the European Synchrotron Radiation Facility and Diamond synchrotrons for access to beamlines. The work was funded by the Royal Society, the Wellcome Trust and the European Research Council (M.S.D.), Cancer Research UK (D.B.W.) and EMBO (W.W.K.).

Author Contributions W.W.K. and D.B.W. designed the experiments. W.W.K., X.F., M.W. and N.B.C. performed the experiments. W.W.K., M.W., M.S.D. and D.B.W. analysed the data and prepared the manuscript.

Author Information The coordinates have been deposited at the PDB with accession codes 4CEH, 4CEI and 4CEJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B.W. (Dale.Wigley@icr.ac.uk).

©2014 Macmillan Publishers Limited. All rights reserved

CORRIGENDUM

doi:10.1038/nature13170

Corrigendum: *Ptpn11* deletion in a novel progenitor causes metachondromatosis by inducing hedgehog signalling

Wentian Yang, Jianguo Wang, Douglas C. Moore, Haipei Liang, Mark Dooner, Qian Wu, Richard Terek, Qian Chen, Michael G. Ehrlich, Peter J. Quesenberry & Benjamin G. Neel

Nature **499,** 491–495 (2013); doi:10.1038/nature12396

After publication of this Letter, we became aware that we had not reported the details of construction and validation of our floxed conditional deletion *Ptpn11* allele. These details are presented in the Supplementary Methods and Supplementary Fig. 1 of this Corrigendum. We also stated in our manuscript that *PTPN11* is not mutated in human hepatocellular carcinoma, but it has been brought to our attention that a low frequency of copy number abnormalities involving the *PTPN11* locus in this disorder has been reported[1]. However, comparable numbers of amplifications and deletions were observed, making it difficult to conclude from the data in ref. 1 that *PTPN11* acts as a tumour suppressor gene in hepatocellular carcinoma.

Supplementary Information is available in the online version of the Corrigendum.

1. Nalesnik, M. A. *et al.* Gene deletions and amplifications in human hepatocellular carcinomas: correlation with hepatocyte growth regulation. *Am. J. Pathol.* **180,** 1495–1508 (2012).

©2014 Macmillan Publishers Limited. All rights reserved

CORRIGENDUM

doi:10.1038/nature13190

Corrigendum: Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia

Liran I. Shlush, Sasan Zandi, Amanda Mitchell, Weihsu Claire Chen, Joseph M. Brandwein, Vikas Gupta, James A. Kennedy, Aaron D. Schimmer, Andre C. Schuh, Karen W. Yee, Jessica L. McLeod, Monica Doedens, Jessie J. F. Medeiros, Rene Marke, Hyeoung Joon Kim, Kwon Lee, John D. McPherson, Thomas J. Hudson, The HALT Pan-Leukemia Gene Panel Consortium, Andrew M. K. Brown, Fouad Yousif, Quang M. Trinh, Lincoln D. Stein, Mark D. Minden, Jean C. Y. Wang & John E. Dick

Nature **506,** 328–333 (2014); doi:10.1038/nature13038

Author Fouad Yousif (of the Ontario Institute for Cancer Research, Toronto, Canada) should have been included in the author list after Andrew M. K. Brown with affiliation number 7 and listed in the Author Contributions as performing and analysing targeted sequencing; these omissions have been corrected in the online versions of this Article. In addition, in the legend to Fig. 1a, "Somatic mutations in *DNMT3a* (*, R882H; †, R137C)" should read "Somatic mutations in *DNMT3a* (*, R882H; †, R326C)".

©2014 Macmillan Publishers Limited. All rights reserved

CORRIGENDUM

doi:10.1038/nature13201

Corrigendum: Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB

Vincent J. Lynch, Gemma May & Günter P. Wagner

Nature **480,** 383–386 (2011); doi:10.1038/nature10595

We inadvertently included a duplicate gel image in the lower right panel of Fig. 3c of this Letter. The gel image should show the loading control ('Input') for the image directly above it (lanes pHsa and dHsa) but instead is a duplicate of the image shown in the upper left panel of Fig. 3c (lanes pAncEu-S5 and dAncEu). Inspection of the original file revealed that the duplicate gel image was superimposed over the correct gel image. Figure 1 of this Corrigendum shows the corrected Fig. 3c. This error does not alter our results or conclusions.

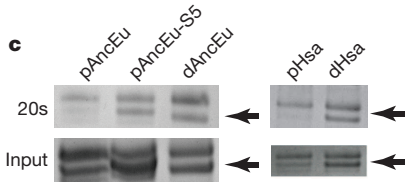


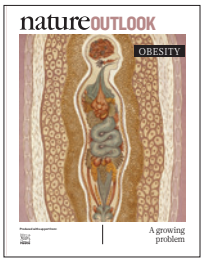
Figure 1 | This is the corrected Fig. 3c.

©2014 Macmillan Publishers Limited. All rights reserved

nature OUTLOOK

OBESITY

17 April 2014 / Vol 508 / Issue No 7496



Cover art: Katie Scott

Editorial
Herb Brody, Michelle Grayson, Tony Scully, Nick Haines, Afsaneh Gray, Vicki Kitchener

Art & Design
Wes Fernandes, Alisdair Macdonald, Lucy Verey

Production
Donald McDonald, Susan Gray, Chris Clough

Sponsorship
Yvette Smith, Reya Silao

Marketing
Elena Woodstock, Hannah Phipps

Project Manager
Christian Manco

Art Director
Kelly Buckheit Krause

Publisher
Richard Hughes

Magazine Editor
Rosie Mestel

Editor-in-Chief
Philip Campbell

For a condition as prevalent and dangerous as obesity (see page S50), we know surprisingly little about its causes and cures. We have much to learn about how fat tissue stores and burns lipids; there may even be new types of human fat cell yet to be discovered (S52). And although it is clear that the types of microbe living in the gut correlate with body weight, we do not know whether changes in these populations are a cause of weight gain, or a consequence (S61).

The best way to lose weight is to eat less and exercise more. But as a strategy to combat obesity at the population level, this common-sense prescription is proving ineffective over the long term. Tailored treatment programmes that factor in the stresses and temptations of the real world, using insights from behavioural research, are showing some success. Drugs may also form part of the solution (S54). Or perhaps the pharmaceutical option should be a last resort, and society should instead use the power of government regulation to encourage healthier lifestyle options (S57).

Of course, obesity does not result from the environment alone — it is one of our most strongly genetically influenced traits. Scores of genes have been implicated, but the evidence suggests that something other than genes accounts for whether someone is likely to become obese (S58).

Controlling appetite is not just a matter of will power; much of our dietary behaviour is hardwired. Neuroscientists are using new techniques to map the neural circuits that control when and how much we eat (S64). But these appetite systems, which evolved to ensure we have enough of the right nutrients, are now being subverted by modern food processing (S66).

We are pleased to acknowledge the financial support of Nestlé Research Center in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

Tony Scully
Science Editor, Nature Outlook

CONTENTS

S50 PUBLIC HEALTH
Society at large
Charting the rise and rise of obesity

S52 CELL PHYSIOLOGY
The changing colour of fat
Beige is the new brown

S54 TREATMENT
Marginal gains
What can you do to keep the weight off?

S57 PERSPECTIVE
Obesity is not a disease
D. L. Katz argues that popping pills isn't the answer

S58 HERITABILITY
The family roots of obesity
Parents pass on epigenetic changes

S61 MICROBIOME
A complicated relationship status
What's the link between body fat and gut bacteria?

S64 NEUROSCIENCE
Dissecting appetite
Mapping the neural circuits that control our hunger

S66 PERSPECTIVE
The tricks of the trade
Processed food can fool the appetite, say Stephen J. Simpson and David Raubenheimer

COLLECTION

S67 **A small-molecule AdipoR agonist for type 2 diabetes and short life in obesity**
Miki Okada-Iwabu et al.

S74 **An excitatory paraventricular nucleus to AgRP neuron circuit that drives hunger**
Michael J. Krashes et al

S79 **Obesity, Food, and Addiction: Emerging Neuroscience and Clinical and Public Health Implications**
Marc N. Potenza

S81 **EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex**
Haruya Ohno et al.

S86 **Diet rapidly and reproducibly alters the human gut microbiome**
Lawrence A. David et al.

S91 **Obesity-associated variants within *FTO* form long-range functional connections with *IRX3***
Scott Smemo et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, Nature **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2014). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Obesity* supplement can be found at http://www.nature.com/nature/outlook/obesity

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

Copyright © 2014 Nature Publishing Group

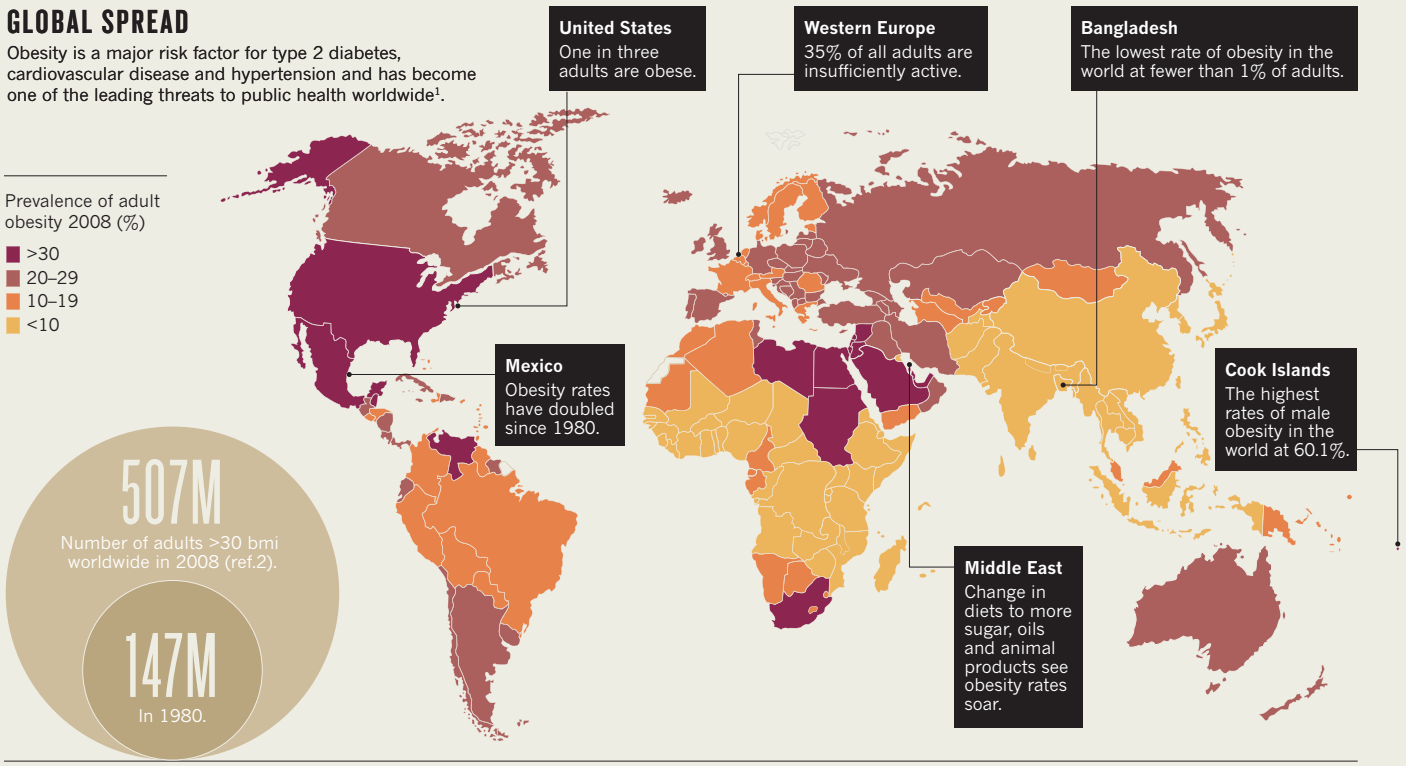
© 2014 Macmillan Publishers Limited. All rights reserved

SOCIETY AT LARGE

The increasing prevalence of obesity is a worldwide phenomenon, affecting peoples from diverse cultural and economic backgrounds. By **Tony Scully**.

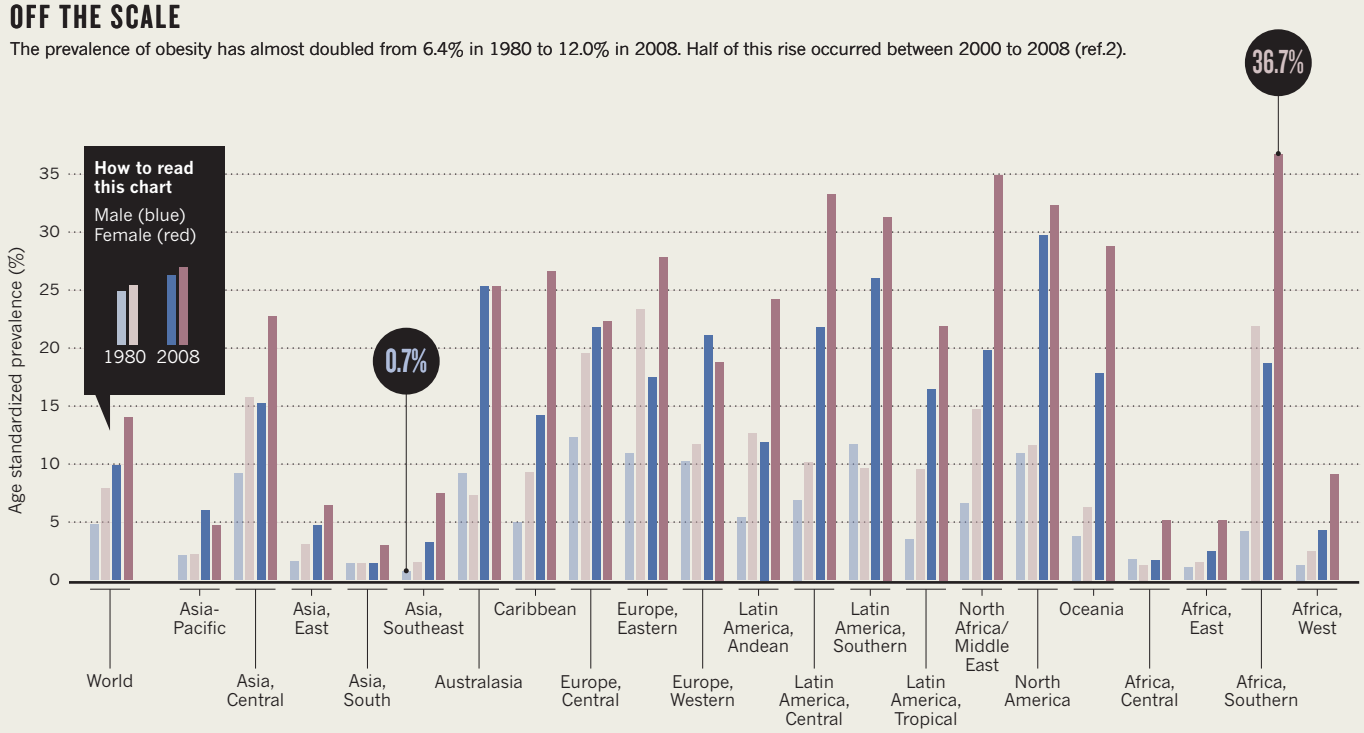
GLOBAL SPREAD

Obesity is a major risk factor for type 2 diabetes, cardiovascular disease and hypertension and has become one of the leading threats to public health worldwide[1].



OFF THE SCALE

The prevalence of obesity has almost doubled from 6.4% in 1980 to 12.0% in 2008. Half of this rise occurred between 2000 to 2008 (ref.2).



© 2014 Macmillan Publishers Limited. All rights reserved

BIGGER BUDGETS

Increasing urbanization can negatively affect diets. People tend to eat more energy-dense foods containing more fat and sugar while reducing their intake of complex carbohydrates and dietary fibre[3].



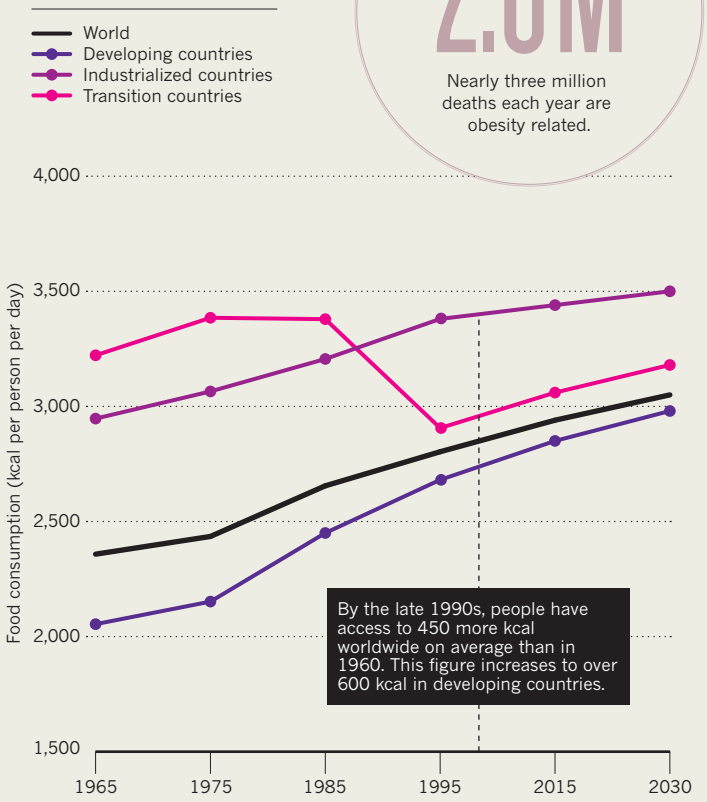
In low- and lower-middle income countries, obesity among women was approximately double that among men.

Only 7% of adults in lower-middle-income countries are obese, this jumps to 24% in upper-middle-income countries.

Many more women than men are obese, with the exception of high-income countries where levels are similar.

RISE OF THE CALORIES

Since the 1960s, the availability of calories per person has risen steadily across all economic zones[4].

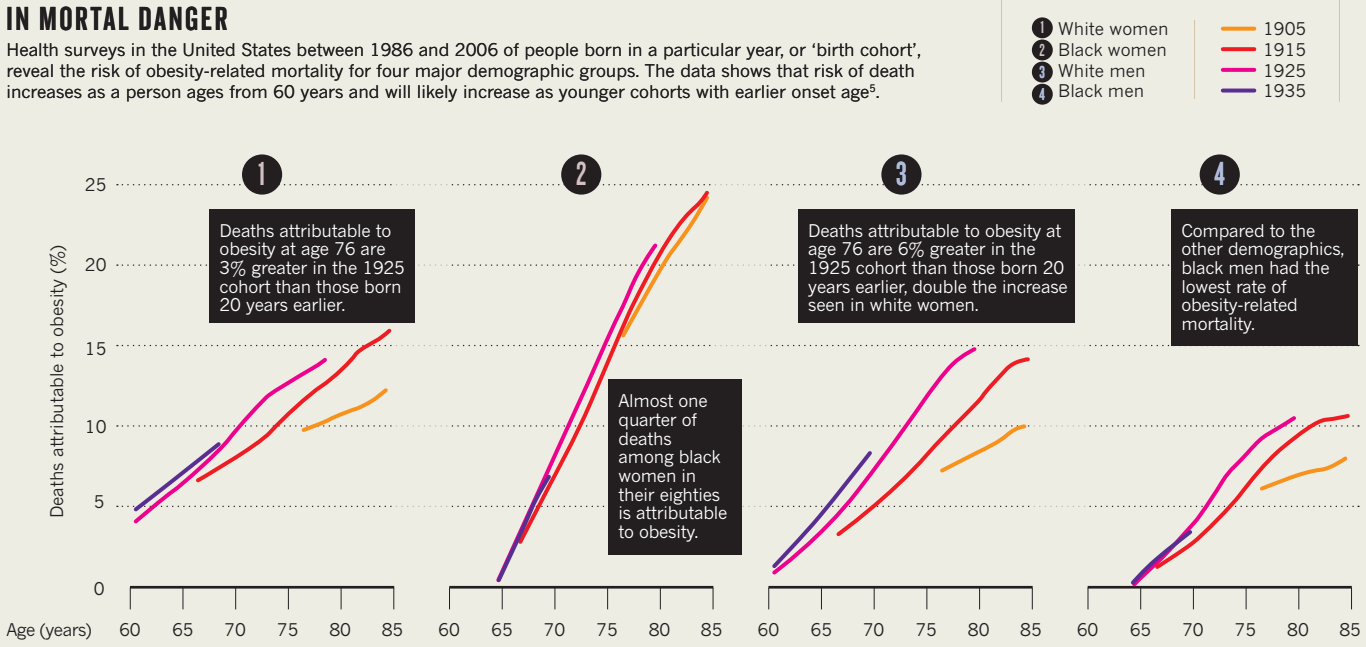


2.8M

Nearly three million deaths each year are obesity related.

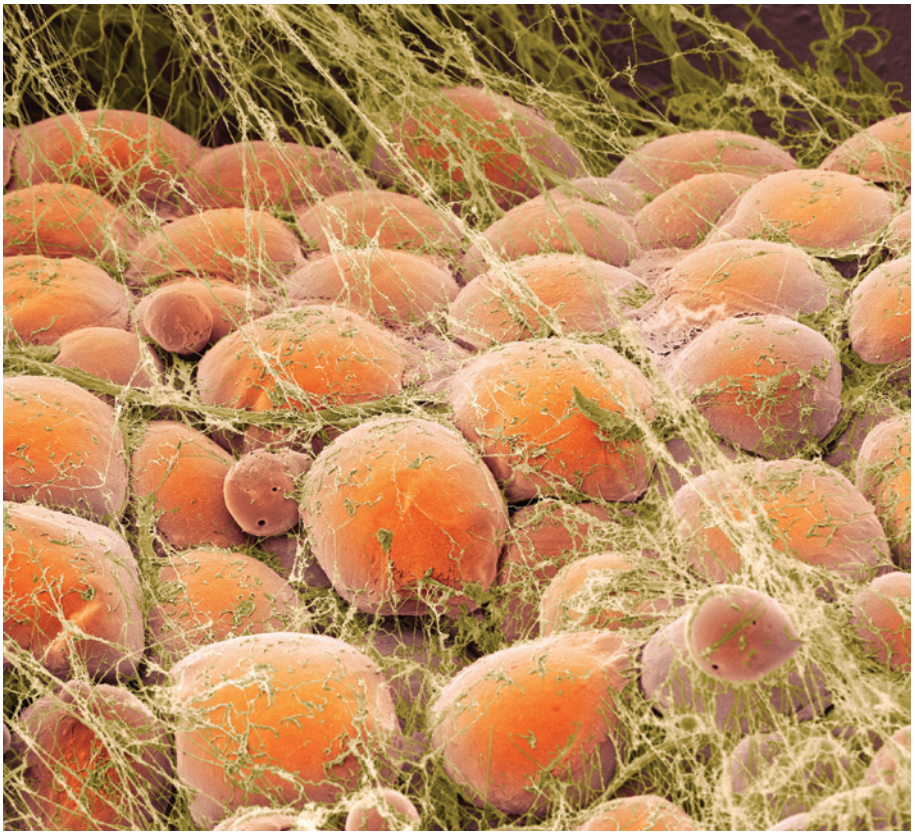
IN MORTAL DANGER

Health surveys in the United States between 1986 and 2006 of people born in a particular year, or 'birth cohort', reveal the risk of obesity-related mortality for four major demographic groups. The data shows that risk of death increases as a person ages from 60 years and will likely increase as younger cohorts with earlier onset age[5].



References: 1. The WHO Global InfoBase.; 2. Stevens, G. A. *et al. Population Health Metrics* **10**, 22 (2012).; 3. WHO Global Health Observatory.; 4. *World agriculture: towards 2015/2030 (Summary report.* Rome, Food and Agriculture Organization of the United Nations, 2002).; 5. Masters, R. K. *et al. American Journal of Public Health* **103**, 1895–1901 (2013).

© 2014 Macmillan Publishers Limited. All rights reserved



STEVE GSCHMEISSNER/GETTY IMAGES

Researchers are attempting to harness the ability of brown cells to burn fat.

CELL PHYSIOLOGY

The changing colour of fat

The different functions of white, brown and beige fat might yield new targets in the fight against obesity and metabolic disease.

BY BRIAN OWENS

When you think of fat in the human body, you might picture a homogenous, white substance, much like a block of lard. But researchers are learning that the role of fat in metabolism changes depending on where it is in the body, and even on the type of fat cell. Soon these differences could be harnessed to fight metabolic disorders such as diabetes and obesity.

The predominant form of fat in mammals, including in humans, is known as white fat. White fat cells are found throughout the body, primarily under the skin, as well as in larger deposits in the abdomen. They are highly specialized for energy storage, hoarding calories in the form of lipid droplets.

The white cells are complemented by brown fat cells, which specialize in converting chemical energy from glucose and lipids into heat, generally when an animal is exposed to the cold. This burning of carbohydrates and fat also helps counteract obesity and metabolic disorders.

But, says Patrick Seale, a developmental biologist at the University of Pennsylvania in Philadelphia, that does not mean the two types of fat are heroes or villains. "Often one portrays white adipocytes as the bad guy, and the brown adipocytes as the good fat cell," he says. "That's not really the case — white fat cells serve an important role in protecting against metabolic disease, but the key is to keep those cells healthy."

NATURE.COM
Brown and beige fat cells and their therapeutic potential:
go.nature.com/nnzf1n

Healthy white fat protects the body by providing a 'safe home' for lipids, which can be toxic to other tissues such as muscle or the liver. So these fat cells hold on to the lipids until the energy they are storing is needed, when they release them into the blood.

Where white fat cells are located also matters. Subcutaneous white fat, scattered throughout the body in depots under the skin, is not associated with metabolic disorders, and may actually be preventative. It is only the abdominal white fat that, when it gets out of control, can cause health problems. And, adds Seale, people with too few fat cells, such as those with a condition called lipodystrophy, have many of the same sorts of problems as people with too much fat. "When they're functioning appropriately, white fat cells will limit metabolic disease," he says. "They're actually a protective cell type."

"When functioning appropriately, white fat cells will limit metabolic disease."

Problems arise when white fat cells store too much lipid. They begin expanding and proliferating rapidly in a process that resembles the growth of a solid tumour, says Philipp Scherer, a cell biologist at the University of Texas Southwestern in Dallas. The blood supply cannot keep up with this expansion, and the cells begin to suffer from lack of oxygen. This hypoxia attracts the protein HIF-1α, which in fat tissue stimulates the extracellular matrix surrounding the cells, leading to fibrosis. The huge, oxygen-starved fat cells do not have enough room to expand and get squeezed to death, releasing their lipid cargo.

As the cells start to die, macrophages swarm to the fat depots to try and clean up the mess by carrying away the lipid droplets. The problem is that the macrophages cannot clear up the lipids fast enough, and so they begin to spill over into other tissues, such as the liver and pancreas. "Those tissues form lipid droplets as well, but they aren't quite as good at it as fat tissue," says Scherer. This inappropriate accumulation of lipids in fat depots and in tissues that are not equipped to handle them results in continued low-grade inflammation and, ultimately, in metabolic disease, insulin resistance and type 2 diabetes.

Until a few years ago, it was generally thought that after infancy humans had only white fat, in contrast to hibernating animals and small rodents, in which brown fat is retained. "In babies it was considered some sort of strange vestige, a little like an appendix," says Jan Nedergaard, a physiologist at Stockholm University. "So it did not attract much attention outside a small circle."

That all changed in 2007, says Nedergaard, when he and his colleagues studied human

© 2014 Macmillan Publishers Limited. All rights reserved

adults and found brown fat deposits around the neck and collarbone — similar to where brown fat is found in mice[1]. "That changed the whole interest in brown fat," he says. "It became a tissue that could be of relevance to controlling body weight and other things."

NOT JUST BROWN AND WHITE

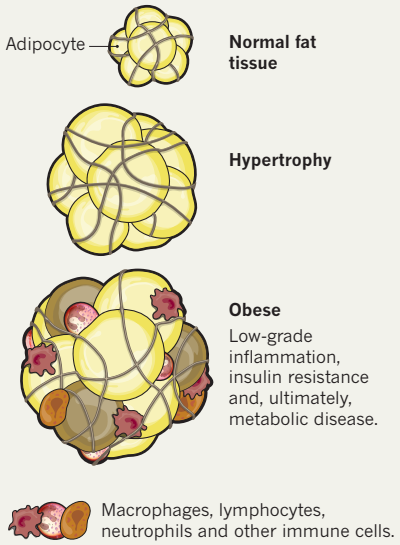
But brown fat is not the end of the story. Scientists had long known that when mice were exposed to cold, brown-fat-like cells would start to crop up in their white fat and begin producing heat by burning lipids, glucose and any other energy-storing molecules available to them. This is not a case of white cells suddenly transforming themselves into brown cells, says Bruce Spiegelman, a cell biologist at Harvard University in Cambridge, Massachusetts. Rather, the white fat depots are seeded with immature cells, which begin to change when exposed to the cold. But these cells are also not exactly brown fat, either. Although they seem to have similar functions, and express many of the same genes as brown fat, they are sufficiently distinct to be considered a new type of fat cell[2], says Spiegelman. Sticking with the colour theme, he calls them beige fat cells. And it is not just mice that have this third type of fat. In 2012, Spiegelman showed that most of the brown fat in humans is actually beige.

There is some disagreement, however, about the relative importance to metabolism of the two types of lipid-burning cell. Although Nedergaard agrees that humans have more beige than brown fat, he thinks that, taken all together, beige cells are less able to burn lipids and produce heat than brown cells. The beige fat cells "clearly have a lower capacity all in all," he says. "We think the main burning is in classical brown fat."

Nedergaard is also not convinced that brown and beige fat have different roles in the body, and so is doubtful that beige fat is of any particular interest for treating human disease. But Spiegelman believes that a study he published earlier this year should put those doubts to rest. Spiegelman's team selectively disabled the beige fat in mice by targeting the protein PRDM16, which is found only in these cells, leaving the white and brown fat intact. The animals ended up with severe metabolic dysfunction — obesity, insulin resistance and damaging fat accumulation in the liver, known as hepatic steatosis[3]. The loss of the beige fat destroys the protective abilities of subcutaneous white fat. "Nobody can look at the data and say the animals aren't screwed up when they lose their beige fat," says Spiegelman. "I think we have to consider the discussion closed in terms of whether beige fat is important in normal physiology."

TISSUE TURNS BAD

Cells overloaded with lipids rapidly outgrow oxygen supply, leading to cell death and inflammation.



FUELLING INTEREST

Whether the target is brown or beige, fat tissue is attracting attention as a possible way to fight obesity and metabolic disease. Researchers are hoping to stimulate the growth of more of the helpful fat cells. "If you have enough of them, and they're efficiently activated, they can alter the energy balance of the body," Seale says.

The most natural way to do that is through cold exposure. "If you said you wanted to build your beige fat, I'd say go lightly dressed in a Canadian winter for a week or two and you'll have plenty of beige and brown fat," says Spiegelman. However, although some researchers have suggested that turning down the thermostat might be a good way to fight obesity[4], any effect will last only as long as the drop in temperature, and exposure to the cold is not really a practical therapy. It would be better, says Seale, if researchers could find a way to stimulate the biochemical pathway that activates the cold response.

One receptor involved in this pathway is known — the β3 adrenergic receptor. In mice, administering a compound that activates it (an agonist) prevents obesity and lowers the incidence of diabetes[5]. Unfortunately, says Seale, no equivalent agonist has been found in humans and the search has largely been abandoned by pharmaceutical companies.

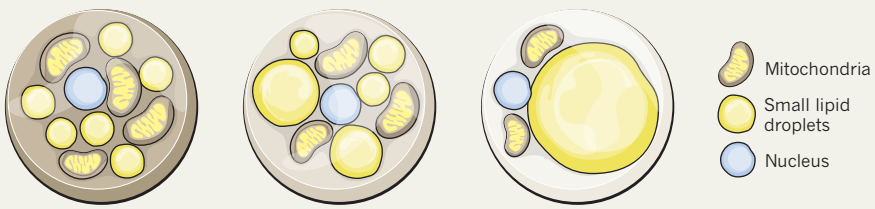
There are other ways to stimulate the activity of brown and beige fat, however. Physical activity seems to do it, says Spiegelman. Irisin, a hormone produced by skeletal muscle in response to exercise, can induce the formation of more beige fat[6]. The hormone FGF21 is another promising candidate. Obese humans treated with a synthetic version of FGF21 showed improvements in their cholesterol counts and body weight[7]. "It's only a matter of time before some of these things become targets for chemicals that can activate the pathways," says Seale.

In the meantime, there is still much to be learned about the characteristics of the different types of fat cells themselves, including the question of whether there are any additional categories waiting to be discovered. One possibility, says Seale, is that white subcutaneous and white abdominal fat — which have very different effects on metabolism — might actually be distinct. "If you take the cells out and isolate them, they have different gene-expression profiles and they really just look like different cell types," he says. "Fat cells might behave very differently and respond differently to different diets and disease states."

Spiegelman agrees. He likens the present understanding of fat cells to changing concepts of white blood cells. "It became B cells and T cells, and monocytes, and macrophages. And macrophages became dendritic cells, and T cells became regulatory T cells and killer T cells," he says. "It's entirely possible that within the category of white we have two or three different categories, and within the category of beige maybe we'll have two or three different categories." ■

SHADES OF A CELL

Three distinct types of fat cell have been identified to date, are there more to follow?



Brown fat cell
Converts chemical energy to heat to protect against cold weather.

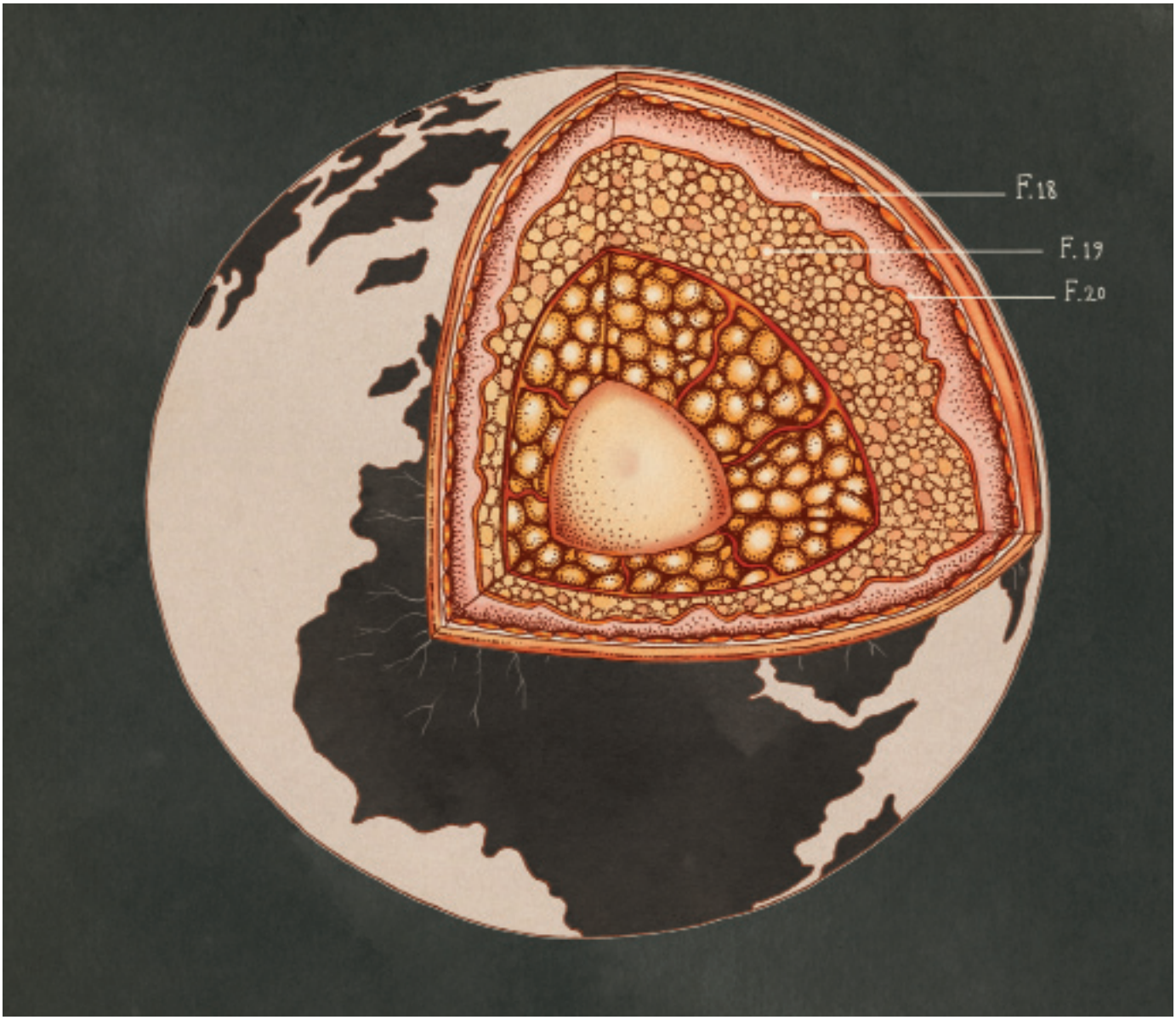
Beige fat cell
Immature cell in white fat tissue matures to burn fat.

White fat cell
Most common fat cell, used to store fat and found beneath the skin and abdomen.

Brian Owens *is a freelance science writer based in St Stephen, New Brunswick, Canada.*

1. Nedergaard, J., Bengtsson, T. & Cannon, B. *Am. J. Physiol. Endocrinol. Metab.* **293**, E444–E452 (2007).
2. Wu, J. *et al. Cell* **150**, 366–376 (2012).
3. Cohen, P. *et al. Cell* **156**, 304–316 (2014).
4. van Marken Lichtenbelt, W. *et al. Trends Endocrinol. Metab.* **25**, 165–167 (2014).
5. Ferrer-Lorente, R. *et al. Life Sci.* **77**, 2051–2058 (2005).
6. Boström, P. *et al. Nature* **481**, 463–468 (2012).
7. Gaich, G. *et al. Cell Metab.* **18**, 333–340 (2013).

© 2014 Macmillan Publishers Limited. All rights reserved



KATIE SCOTT

TREATMENT

Marginal gains

Behavioural interventions work, but not for everyone, and weight regain is common. Are there better ways to treat obesity?

BY EMILY ANTHES

In Providence, Rhode Island, researchers are teaching overweight adults how to get more sleep. The patients are learning the importance of developing regular night-time routines, avoiding alcohol and caffeine before bed, and other basic 'sleep hygiene' strategies as part of a study investigating whether getting more shut-eye can lead to healthier eating habits and weight loss. In San Francisco, California, clinicians are sending overweight, low-income, pregnant women to a course in mindfulness, in the hope of reducing stress-related overeating. In New York City, scientists are asking overweight African-American and Latino adults to make one small change in their eating behaviours, such as using smaller plates.

These small-scale trials are all part of Obesity Related Behavioral Intervention Trials (ORBIT), a US$37-million programme at the US National Institutes of Health, based in Bethesda, Maryland, designed to target some of the major drivers of weight gain and to translate insights from basic behavioural science research into innovative treatments for obesity.

Despite years of research and countless clinical trials, there is no clear strategy for treating obesity at the population level and reversing the

NATURE.COM
A good night's sleep helps to regulate metabolism:
go.nature.com/mropgx

© 2014 Macmillan Publishers Limited. All rights reserved

obesity epidemic. Comprehensive behavioural programmes, designed to get patients to eat less and exercise more, can yield significant weight loss. But they do not help everyone. Certain segments of the population, such as those with low incomes, remain especially difficult to reach, and many patients have trouble maintaining their hard-earned weight loss once a study ends.

"We need to work on ways to help people initiate these changes in their lives and stick to them."

"We have developed some good interventions for weight control, and those can be very effective in helping people achieve clinically significant weight loss," says Susan Czajkowski, a research psychologist at the National Heart, Lung, and Blood Institute in Bethesda, Maryland, and the lead project officer of ORBIT. "But we need to work on ways to help people initiate these changes in their lives and to stick with them."

TAKE A LOAD OFF

The first-line treatment for obesity is getting a patient to change his or her behaviour: eat fewer calories and exercise more. If obese and overweight patients lose 5–10% of their body weight, they show significant clinical improvements, including reductions in various cardiovascular risk factors. Over the years, studies have shown that lifestyle changes can yield this amount of weight reduction, at least in the short term.

In late 2013, the American College of Cardiology, based in Washington DC, the American Heart Association in Dallas, Texas, and The Obesity Society in Silver Spring, Maryland, jointly issued a new set of guidelines for treating overweight and obese adults based on an evaluation of peer-reviewed research by an expert panel[1]. Among other treatment strategies, the panel looked at the effectiveness of comprehensive behavioural interventions that included three components: a low-calorie diet, increased physical activity, and some sort of structured behavioural programme in which patients learn strategies that can help facilitate weight loss (such as goal setting or regular monitoring of calorie intake and activity). The programmes also tended to include weekly meetings with health professionals.

After 6 months, such interventions produce an average weight loss of up to 8 kg, a figure that represents a 5–10% weight loss for many patients. But participants' weight tends to level off at 6 months. And the panel found that after a year, even with continued regular contact with professionals, patients begin to put the weight back on, at an average rate of 1–2 kg per year. "We can get a lot of weight loss," says Catherine Loria, a nutritional epidemiologist at the National Heart, Lung, and Blood Institute and a co-author of the new guidelines. "The main challenge there is still keeping the weight off long term."

CREEPING BACK

The slow regain of body fat, which scientists have documented again and again, reveals how difficult it can be to change long-term habits. New skills that people learn in behavioural programmes can easily break down in the face of real-world environments that are saturated with calorie-dense foods. "We've got these sort of tepid behavioural interventions — 'Do this diet for a little while' or 'Learn these few skills' — but meanwhile we have a culture that sanctions junk as a food group," says David Katz, director of the Yale-Griffin Prevention Research Center in Derby, Connecticut. "The modern world conspires against weight control."

Helping people maintain their weight loss in a society with an abundance of cheap convenience foods and oversized portions may require long-term, ongoing support. In a study of more than 1,000 people who had lost an average of 8.5 kg in 6 months, researchers found that individual monthly counselling sessions — which were conducted primarily by telephone and gave patients a chance to report on their progress, set future goals and discuss challenges they faced — helped participants maintain some of their weight loss[2]. At the 3-year mark, participants who had personal contact with counsellors had regained 4 kg, on average, whereas those who participated in an online maintenance programme had put 5.2 kg back on, and patients receiving neither treatment regained 5.5 kg. "We've started to think about it as a chronic disease which needs continual intervention," Loria, who co-authored the study, says of obesity. "The more involvement you have, the more contact you have, probably the better a person will do long term."

Comprehensive behavioural programmes have also shown promise in children. A 2012 review of randomized controlled trials of lifestyle interventions for obese children concluded that such treatment programmes led to significant reductions in body-mass index (BMI) — a standard ratio of weight to height that is commonly used to determine whether people are overweight or obese — and improvements in cholesterol levels, glucose metabolism and blood pressure[3].

Particularly for school-age children, "we have a treatment approach that can produce effective changes," says Chantelle Hart, a clinical psychologist at the Center for Obesity Research and Education at Temple University in Philadelphia, Pennsylvania. However, she adds, "Not all children who participate in behavioural interventions achieve a healthy weight status, so we need to look at innovative strategies that we can add as an adjunct."

Hart is developing one such approach. Epidemiological studies have revealed a connection between sleep habits and body weight. Children and adults who sleep less tend to weigh more, and sleep deprivation can alter the levels of hormones involved in metabolism and appetite. Last year, Hart published the results of the first experimental study on sleep duration and weight in school-age children. She found that when these children were assigned to increase their sleep duration — spending an extra hour-and-a-half in bed — they ate 134 fewer calories a day, on average, than when they were asked to spend less time than usual in bed[4]. After a week on the enhanced sleep schedule, the children had lost an average of 0.22 kg.

At the University of Buffalo in New York, clinical psychologist Leonard Epstein is developing a programme to enhance the ability of overweight children and adults to delay gratification. "So far, at least in the laboratory, we've been very successful," he says. In a paper published in 2013, he showed that overweight women who were asked to think about positive events in their future — such as vacations, birthdays or holiday celebrations — consumed fewer calories during a 15-minute snacking session[5].

Obesity is increasingly an issue of economic class. Although obesity rates have begun decreasing among adolescents of high socioeconomic status, no such improvement is evident in poorer and minority populations. In March 2014, headlines trumpeted a study finding that the proportion of 2- to 5-year-old children who are obese in the United States dropped from 13.9% in 2003–2004 to 8.4% in 2011–2012 (ref. 6). However, the data also showed that obesity remained more common among the Hispanic and African-American children in that age group.

The prevalence of obesity among those without financial resources means that we need inexpensive treatments. "What we do to intervene in that group has to be low cost and has to be something that's attainable and feasible in a lower household income," says Elsie Taveras, a paediatrician and obesity researcher at the Massachusetts General Hospital in Boston.

In a randomized trial published in 2013, Taveras showed that she could reduce

MEASURING WEIGHTS

How do we define obesity?

Obesity is typically assessed by using the body-mass index (BMI), a standard ratio of weight to height. The BMI is calculated by dividing someone's weight, in kilograms, by the square of his or her height, in metres. Patients who have a BMI between 25.0 and 29.9 are considered to be overweight, and a BMI of 30.0 or higher indicates obesity.

© 2014 Macmillan Publishers Limited. All rights reserved

children's BMIs by helping their families make small, no-cost changes in basic home routines[7]. In her "Healthy Habits, Happy Homes" study, Taveras and her colleagues dispatched health coaches to work with low-income families with children between the ages of 2 and 5. (Approximately 60% of the families had annual incomes of US$20,000 or less.) The coaches encouraged families to eat more meals together, reduce television watching and increase children's sleep duration — all behaviours that have been linked with reduced obesity risk. After 6 months, children in families that had received the intervention were sleeping longer and spending less time watching TV; their BMIs had also decreased by an average of 0.18, whereas those of children in the control group had increased by 0.21, on average.

Different patient populations have different needs, and obesity researchers are moving away from the 'one size fits all' idea, says Czajkowski. ORBIT is funding trials focusing on groups ranging from African-American teens to menopausal women. "We've learned that there are different motivations, different challenges, different barriers for different groups of people, and so by focusing in on subgroups and their unique needs, you get a better chance for success," Czajkowski says.

DRUG DIETS

Some experts believe that the marginal gains seen with behavioural interventions indicate that the only practical solution to the obesity problem will involve drugs. The use of anti-obesity drugs peaked in 1996, when approximately 8,500 prescriptions for the medications were filled for every 100,000 people in the United States. Then, in 1997, two popular medications, the drugs fenfluramine and dexfenfluramine, were linked to cardiovascular problems and removed from the market. In 2011, there were about 2,500 prescriptions filled per 100,000 people and 2.74 million total users of anti-obesity drugs[8].

At that time, there was just one drug available that the US Food and Drug Administration (FDA) had approved for long-term use against obesity: orlistat (Xenical), which interferes with the body's ability to absorb fat. When taken at high doses, and in conjunction with a low-calorie diet and increased exercise, orlistat yields weight loss of about 3% beyond what patients achieve with a placebo, as well as an improvement in several cardiovascular risk factors and a reduction in the risk of diabetes[9]. In 2012, the FDA approved two drugs for the long-term treatment of obesity: lorcaserin (Belviq), which promotes satiety (the feeling of being full) by activating a certain kind of serotonin receptor in the brain, and a phentermine/topiramate combination (Qsymia), which combines an amphetamine that suppresses appetite with an anti-epileptic drug associated with weight loss (see 'Weighing up the options').

"There's no question that there are many people who lose more weight when they take medication."

Like most drugs, however, these medications can sometimes have serious side effects. Qsymia, for instance, can cause an elevated heart rate and, when taken during pregnancy, birth defects. And over the past two decades, six weight-loss compounds have been withdrawn from the market due to safety concerns, most recently in 2010 when sibutramine (Meridia) was pulled after research linked the drug to an increased risk of heart attacks and strokes.

WEIGHING UP THE OPTIONS

Weight-loss drugs approved by the US Food and Drug Administration for long-term use.

Drug	Approved	Average placebo-adjusted weight loss	Percentage losing at least 5% of body weight
Xenical (orlistat)	1999	3%	57%, compared with 31% of those taking a placebo
Belviq (lorcaserin)	2012	3.0–3.7%	47%, among patients without type 2 diabetes, compared with 23% of those taking a placebo
Qsymia (phentermine and topiramate)	2012	6.7–8.9%, depending on the dose	62–69%, depending on the dose, compared with 20% of those taking a placebo



Pregnant mothers participated in a study to reduce stress-related overeating.

MATERNAL ADIPOSITY, METABOLISM AND STRESS (MAMAS) STUDY

Such side effects may be unavoidable. The pathways that are involved in metabolism and obesity are so ubiquitous in the body that it is difficult to alter them in ways that lead to weight loss without unwanted consequences, says Tamas Horvath, a neurobiologist at Yale University in New Haven, Connecticut. "You are interfering with autonomic regulation of body tissues, which in the long run may not be beneficial for those tissues or the body at large," he says.

But that does not mean that weight-loss drugs cannot be useful, says Horvath. He believes that regulators need to be more willing to tolerate their side effects, at least in the morbidly obese, for whom the benefits of losing weight may outweigh potential risks.

Obesity drugs can also be useful for kick-starting weight loss and helping patients stick to low-calorie diets. Researchers have found, for instance, that patients who take a weight-loss drug while participating in lifestyle interventions lose 3–9% more weight than those who get a placebo[10]. "There's no question that there are many people who lose more weight when they take medication," says Louis Aronne, director of the Center for Weight Management and Metabolic Clinical Research at New York-Presbyterian Hospital and the Weill Cornell Medical Center.

For now, though, with much left to learn about the long-term effects of these drugs, most experts recommend that patients try diet and exercise first and that weight-loss medications be used only in conjunction with behavioural approaches. As Loria puts it, "I don't know that we can be on weight-loss drugs for 60 years, and so at some point, I think people need to make sustainable changes in what they eat and the activity they get."

Ultimately, the only effective, sustainable solution — on a large scale over the long term — may be to change the culture. "The issue to me is not whether or not behavioural interventions work," Katz says. "There's absolutely no question to me that they are the thing that we have to depend on." But even the most comprehensive programme, he says, is no match for a culture built around calorie-dense foods and sedentary lifestyles. "Behavioural interventions are a very slow march forward on a moving walkway that's going in reverse." ■

Emily Anthes *is a freelance writer based in Brooklyn, New York.*

1. Jensen, M. D. *et al. J. Am. Coll. Cardiol.* http://dx.doi.org/10.1016/j.jacc.2013.11.004 (2013).
2. Svetkey, L. P. *et al. JAMA* **299**, 1139–1148 (2008).
3. Ho, M. *et al. Pediatrics* **130**, e1647–e1671 (2012).
4. Hart, C. N. *et al. Pediatrics* **132**, e1473–e1480 (2013).
5. Daniel, T. O., Stanton, C. M. & Epstein, L. H. *Psychol. Sci.* **24**, 2339–2342 (2013).
6. Ogden, C. L. *et al. JAMA* **311**, 806–814 (2014).
7. Haines, J. *et al. JAMA Pediatr.* **167**, 1072–1079 (2013).
8. Hampp, C., Kang, E. M. & Borders-Hemphill, V. *Pharmacotherapy* **33**, 1299–1307 (2013).
9. Rucker, D. *et al. BMJ* **335**, 1194 (2007).
10. Yanovski, S. Z. & Yanovski, J. A. *JAMA* **311**, 74–86 (2014).

© 2014 Macmillan Publishers Limited. All rights reserved

PERSPECTIVE



Obesity is not a disease

The misguided urge to pathologize this condition reflects society's failure to come to terms with the need for prevention, says **D. L. Katz**.

Doctors are historically ill-prepared to address the problem of obesity — with a tendency either to ignore it, or to ineffectually wag an admonishing finger[1–3]. To focus physicians' attention on this prevalent health issue, the American Medical Association (AMA) recently declared obesity to be a disease. This well-intentioned move by the AMA is misguided in that it implies that tens of millions of people must now have bodies or minds, or both, that are not working properly. Even seemingly healthy, but heavy, people — adults and children alike — are now, by definition, diseased[4]. Imposing such a status has broad ramifications for society and requires careful reflection.

The standard measure of obesity is body-mass index (BMI), which is roughly speaking the ratio of weight to height. For adults, a BMI greater than 30 is associated with increased risk of illness, disability and death. However, a risk factor is not a disease, because each can occur independently of the other.

Obesity is an important contributor to the prevailing burden of chronic disease, lying on the causal pathway to much of what plagues modern society and its people — heart disease and diabetes to name two of the most serious. However, not only can these diseases develop in the absence of obesity, but not everyone with a high BMI develops any such condition.

The categorization of obesity as a disease could have a pernicious influence on efforts to remedy the problem at its true origins. The treatment of diseases customarily involves drugs, medical technology, clinic visits and surgical procedures. If obesity is a disease, the therapeutic advances on which its management depends presumably reside in these domains.

> "THE CATEGORIZATION OF OBESITY AS A DISEASE COULD HAVE A PERNICIOUS INFLUENCE ON EFFORTS TO REMEDY THE CONDITION."

HEAVY COSTS

The disease approach would impose substantial costs. Obesity affects many tens of millions of adults and children in the United States alone[5]. If we were now to conclude that all these individuals warrant disease treatment, the collective need for drugs and bariatric surgery would be staggering. That would mean not only a huge financial outlay, but the imposition of a vast array of side effects on the overweight population. Even the best of drugs are prone to side effects, and to date, weight management has been forced to rely on anything but the best of drugs[6]. The long-term effects of bariatric surgery are still highly uncertain as well — particularly for ever younger candidates. Even if surgery proves sustainably effective, the need to rely on the rearrangement of natural gastrointestinal anatomy as an alternative to better use of feet and forks seems a societal travesty[7].

The consequences of labelling obesity a disease seem to be a price the medical profession is willing to pay to legitimize the condition. It may also be an attempt to own it, and the profits that come with treating it.

Our bodies, physiologies and genes are the same as they ever were. What has changed while obesity has gone from rare to pandemic is not within, but all around us. We are drowning in calories engineered to be irresistible[8]. We are awash in labour-saving technologies and a societal mindset that urges us to use all that we invent[9].

LIFEGUARDS

Like breathing air, our capacity to get fat is part of normal physiology. The dividing line between normal and abnormal fat accumulation has nothing to do with fat accumulation per se, but rather whether or not those energy reserves, a fundamental survival strategy of omnivorous and carnivorous animals, are ever drawn down. In our ancestral context, they were; in our modern context, they never are.

We don't wait for people to drown and devote our focus to resuscitation. Instead, we do everything we can to prevent drowning in the first place: we erect fences around swimming pools, station lifeguards at beaches, offer swimming lessons, and keep a close eye on one another at the water's edge. People still drown, so we need medical intervention as well. But that is a last resort, far less good than prevention, and applied far less commonly. There is an analogous array of approaches to obesity prevention and control. These include environmental reforms, such as making stairs, pavements and bike lanes more readily available, and altering food service settings to encourage the more healthy choices; social reforms, such as making physical activity programming a standard aspect of every work and school day; policy reforms, such as regulation of both food formulations and food marketing; and skill-building, including teaching adults and children how to identify more nutritious foods and how to cook.

Obesity warrants medical as well as cultural legitimacy and respect, but needs not be a disease to earn them. Calling obesity a disease contradicts the functioning of our bodies, and implies a blame residing there. But the blame for hyperendemic obesity, and its best remediation, resides not within bodies that work as they ever did, but all around, with the collective actions of the body politic. ■

D. L. Katz *is the founding director of Yale University's Prevention Research Center, president of the American College of Lifestyle Medicine, and editor-in-chief of the journal* Childhood Obesity.
e-mail: davkatz7@gmail.com

1. Nawaz, H. & Katz, D. L. *Am. J. Prev. Med.* **21**, 73–78 (2001).
2. Nawaz, H. *et al. Am. J. Public Health* **89**, 764–767 (1999).
3. Kraschnewski, J. L. *et al. Med. Care* **51**, 186–192 (2013).
4. Katz, D. L. *Childhood Obesity* **10**, 1–3 (2014).
5. Gill, T. P. *et al. Asia Pac. J. Clin. Nutr.* **8**, 75–81 (1999).
6. Cannon, C. P. & Kumar, A. *Clin. Cornerstone* **9**, 55–68; discussion 69–71 (2009).
7. Katz, D. L. School over scalpels. *US News & World Report* (11 January 2013).
8. Moss, M. The extraordinary science of addictive junk food. *The New York Times Magazine* (20 February 2013).
9. Diamond, J. Best invention; invention is the mother of necessity. *The New York Times* (18 April 1999).

© 2014 Macmillan Publishers Limited. All rights reserved



KATIE SCOTT

HERITABILITY

The family roots of obesity

Scores of genes are implicated in obesity, but they cannot account for a family's predisposition to obesity. Are there other ways parents can influence their children?

BY CASSANDRA WILLYARD

Stephen O'Rahilly and Sadaf Farooqi, genetics researchers at the University of Cambridge, UK, have been on the hunt for the genes that drive obesity for more than 15 years. One of their first big breaks came in 1997, when two severely obese cousins from Pakistan were referred to them for a clinical assessment. The eight-year-old girl weighed 86 kg — as much as a tall man — and the two-year-old boy tipped the scales at 29 kg. No matter how much these children ate, they never felt full.

A quick blood test pointed Farooqi and O'Rahilly to the problem: both children lacked leptin, a hormone that regulates appetite. The scientists found that the cousins had a mutation in the gene responsible for leptin production — called *ob* for 'obese' — which had only recently been identified in mice[1]. The cousins provided the first irrefutable evidence that our genes can lead us to pile on the pounds. Researchers have since implicated dozens more genes.

"Obesity is one of the strongest genetically influenced traits that we have," says O'Rahilly. Classic twin studies in the 1980s and 1990s, which relied on pairs of identical and fraternal twins, suggest that 40–70% of variation in body size is due to genetic factors[2].

Although rare genetic mutations in the leptin gene and elsewhere in the genome can cause extreme obesity, most cases seem to be

© 2014 Macmillan Publishers Limited. All rights reserved

influenced by several more common genetic variants that have much subtler effects on weight. Over the past two decades, researchers have begun to zero in on some of these genes. "We have developed a very large and growing body of data on genes and variants that either cause obesity or likely increase the risk of obesity," says Claude Bouchard, an obesity researcher at the Pennington Biomedical Research Center in Baton Rouge, Louisiana. But the picture is far from complete. While some research groups work to uncover other genes that could help explain the wide variability in body weight, others are beginning to investigate whether epigenetics can account for the heritability of obesity.

GWAS REVOLUTION

No research technique has helped to link more genes to obesity than genome-wide association studies (GWAS). Researchers previously had to make educated guesses about which genes might be involved, and then examine a few variants within that gene to see whether any could be tied to obesity. But in 2005, when researchers published the first GWAS, "everything changed", says Ruth Loos, a genetics expert at Mount Sinai Hospital in New York.

Using GWAS let researchers compare the genomes of thousands of obese individuals with those of thousands of lean people. Rather than looking at the entire genome, the researchers examine a set number of sites, called single-nucleotide polymorphisms (SNPs), where variations typically occur. "You screen the whole genome without any hypothesis," Loos says. Variants that consistently show up among obese individuals but not among lean individuals are predicted to be associated with obesity.

In 2007, for example, a GWAS found that individuals who carry two copies of a common variant of the *FTO* gene weigh on average 3 kg more than those who do not carry any copies[3]. The genetic variant seems to predispose people to consume greater quantities of food. And in 2013, a study showed that men who carry two copies of the *FTO* risk allele felt hungrier and had higher levels of the appetite-stimulating hormone ghrelin after they ate than men who do not carry the risk variant[4].

About 75 variants have been linked to obesity in this way, leading Loos to describe the strategy as "really successful". And new, unpublished research from an international consortium known as Genetic Investigation of Anthropometric Traits (GIANT) adds dozens more to the list. Those results "will double the number of known obesity loci", says Paul Franks, a genetic epidemiologist who studies type 2 diabetes at Lund University in Malmö, Sweden.

NATURE.COM
For the latest research on the genetics of obesity:
go.nature.com/cLhzvb

But GWAS studies aren't perfect. They can lead researchers to important parts of the genome, but it can be difficult to sort out which gene within that region might be the culprit. For example, in March, researchers called into question the well-established link between *FTO* and obesity[5]. They think that the real link might involve *IRX3*, a gene whose on/off switch, called a promoter, lies next to *FTO*.

DARK MATTER

Many scientists had first assumed that the heritability of obesity would be explained by common genetic variants. But that hasn't held true. "People found the common variants with the GWAS, and they found many of them," says Farooqi. "But each variant on its own only had a very subtle effect on weight." Taken together, the variants identified through GWAS explain about 2–4% of variability that is attributable to genetics. So if these common variants don't explain the variation, then what does?

"Some people refer to this as the dark matter of quantitative genetics," says Leibel: "We know it's there, but we haven't been able to see it," says Rudolph Leibel, a geneticist at Columbia University in New York. Part of this 'missing heritability' could be explained by genetic variants that have yet to be discovered. "It may be we need to add hundreds and hundreds of SNPs with very small effect sizes to capture more of the genetic variance," Leibel says. Larger GWAS studies might be able to pick some of these up.

Franks offers another possibility. Obesity depends not just on genetics, but also on environment and the interaction between the two. Current GWAS studies don't take this complexity into account, so they may be missing genetic variants that increase the risk of obesity but only under certain conditions. The roughly 75 variants identified by GWAS, "are not intrinsically good candidates for gene–environment interactions because of the way they've been identified," Franks says.

"Obesity is one of the strongest genetically influenced traits we have."

Some of the strongest evidence for gene–environment interactions comes from research on the risk variants of *FTO*. Soon after these variants were discovered in 2007, several groups began reporting that exercise attenuates this risk. Other studies, however, failed to find a similar relationship. To resolve the issue, more than a hundred researchers launched a meta-analysis that included 45 studies involving more than 218,000 adults and 19,000 children[6]. Not surprisingly, they found that people who carry the susceptibility gene had a higher risk of obesity. However, the researchers also observed that the risk appears to be reduced in people who are physically active. "This is a great example of a very prevalent risk allele for obesity which is very powerfully influenced by the environment," Leibel says.

But there could be another explanation: Farooqi thinks that some of this missing heritability may be accounted for by rarer variants that haven't yet been identified. GWAS do a good job of pinpointing common alleles that occur in more than 5% of the population, but these studies can't pick out variants that occur less frequently. Some of these rare variants might have a larger impact on an individual's body size than the more common variants. Farooqi and her colleagues have already shown that rare variants can drive some cases of extreme obesity, and she thinks that such variants might also explain more common forms of obesity. But as she points out, "no one has really shown that yet."

Common SNPs most often occur between genes in stretches of DNA that don't encode proteins, but Farooqi and O'Rahilly speculate that mutations in protein-encoding genes may have a greater effect. By sequencing only the regions of the genome that encode proteins — a process called whole-exome sequencing — they hope to find new rare variants that can explain more of obesity's heritability and provide new understanding of the molecular pathways that drive the condition. Farooqi and O'Rahilly have put together a cohort of 5,000 people who have been obese since childhood, and through whole-exome sequencing of nearly 1,800 of those individuals, they have already found rare, obesity-related variants in the gene *KSR2* (ref. 7).

Precisely identifying the genes and pathways involved in obesity could lead researchers to new therapies. In the case of the Pakistani cousins, the solution was simple: the children needed leptin. After the researchers began administering the hormone, the kilograms started to melt away[8]. But even when the problem is more complex, there may be ways medications that can help. For example, as many as 6% of severely obese individuals have mutations in a gene called *MC4R*[9], which encodes melanocortin 4 receptor, a protein that helps to regulate appetite. In 2003, Farooqi and her colleagues found that the severity of *MC4R* mutation affects the degree of obesity[10]. "If you've got a bad mutation that stops the receptor from working, you're worse off than if you have a mild mutation," she says. A Boston-based pharmaceutical company called Rhythm has developed a compound that stimulates melanocortin 4 receptor, and the hope is that it will help individuals who have one 'good' copy of the gene to lose weight. Rhythm launched a clinical trial to test the therapy in September 2013.

FAMILY HEIRLOOMS

Humans are especially susceptible to environmental stimuli during embryonic development. And studies suggest that what happens in the womb can cause lasting

© 2014 Macmillan Publishers Limited. All rights reserved

changes in gene expression and influence disease risk even in adulthood, a concept known as fetal programming. This raises the possibility that a mother's experiences during pregnancy — such as malnutrition — can influence the next generation.

In the winter of 1944–45, a German blockade starved the western part of the Netherlands. The resulting famine, known as the *Hongerwinter* ('Hunger Winter'), killed more than 20,000 people. A landmark study published in the 1970s looked at 300,000 recruits in the Dutch army to determine whether their risk of obesity was influenced by their experience in the womb[11]. "They found hugely significant effects," says Robert Waterland, who studies the influence of nutrition on development at Baylor College of Medicine in Houston, Texas. Women who became pregnant towards the end of the famine, and so were malnourished during only the first trimester, had sons with higher rates of obesity at age 19 than sons who were born before or after the famine (all of the army recruits studied were male). However, pregnant women who were already in their second trimester when the famine began had sons with a lower risk of obesity. "That was really one of the first demonstrations of programming," says Margaret Morris, an obesity researcher at the University of New South Wales in Sydney, Australia.

Evidence now suggests that these changes were the result of epigenetics, the chemical tags that can change gene expression without altering the genetic code. When researchers looked at the genes of children conceived toward the end of the *Hongerwinter*, they found differences in the prevalence of methyl groups (the most common epigenetic mark) on a gene called *IGF2*, which encodes a hormone that promotes growth during gestation[12]. Children whose mothers had gone hungry during the early part of their pregnancy had fewer methyl tags on *IGF2* than their same-sex siblings who hadn't weathered the famine. These DNA tags tend to turn genes off. But exactly how diminished methylation on *IGF2* might lead to outcomes such as obesity is not yet fully understood.

The epigenetics of obesity isn't only about the mother — the father's experiences can have an impact too. When researchers in Australia fed male rats a fatty diet, the rats — as expected — put on weight and developed signs of diabetes[13]. But, surprisingly, the weight gain also seemed to affect the rats' daughters: the female offspring had trouble controlling their insulin levels despite being on a normal diet. And a study published in 2013 showed that children with obese fathers had less methylation on a particular region of the *IGF2* gene than children who were born to lean fathers[14].

These studies suggest that parents' experiences can have profound effects on the health of their children. What remains unclear is whether these changes are truly heritable.



WELLCOME LIBRARY, LONDON

The *ob* gene, which is responsible for leptin production, was first identified in obese mice.

CLEAN SLATE

Scientists once thought that epigenetic tags could not be passed from parent to child via sperm or egg cells because any epigenetic marks these cells carry are erased soon after fertilization. But in 2013, a study found that some methyl marks can escape the reprogramming process[15]; this indicates that an epigenetic mark in the egg or sperm could be passed on from generation to generation. "What is special about epigenetics is its ability to maintain a phenotype," says Neil Youngson, an epigenetics expert at the University of New South Wales. "But also we're seeing that maybe it's being maintained beyond that into the next generation."

"Identifying the genes and pathways involved in obesity could lead researchers to new therapies."

To show that an epigenetic change is due to inheritance, however, researchers have to look at multiple generations. When a pregnant woman experiences malnutrition or some other environmental stress, three are directly exposed — the mother, the child in her womb, and also her grandchildren because the fetus already contains cells that will eventually become gametes. "We call those programming effects," Youngson says. "But true transgenerational epigenetic inheritance occurs when none of those generations directly experienced the environmental change." So, if a woman's great grandchildren show a particular epigenetic change that is linked to environmental stress during their great grandmother's pregnancy, that change can be said to be inherited rather than being a programming effect. On the father's side, the exposure isn't as far reaching: a fatty diet might change the epigenetics of a man's sperm, but the programming effect will only be experienced by the child and not the grandchild, unless the change is inherited. "While the son of a starved or obese man could be said to have directly experienced the effect," Youngson says, "the grandson did not directly."

There is some limited evidence to support transgenerational epigenetic inheritance of obesity risk, at least in rats. Michael Skinner, who studies epigenetics at Washington State University in Pullman, has exposed pregnant rats to various environmental contaminants: most recently, the pesticide dichlorodiphenyltrichloroethane (DDT)[16]. Although the chemical had no effect on the weight of the rats' offspring, more than half of the rats' great grandchildren grew fat. And Skinner observed epigenetic changes in the rats' sperm. The findings don't necessarily translate into humans, but Skinner argues that "we have to understand that the environment we live in is driving our health and future generations' health."

Although this study and others suggest that epigenetic marks can be passed down through the generations, it remains unclear what role this type of inheritance might have in human obesity — or whether it can explain any of the missing heritability that troubles geneticists.

Researchers working to untangle how genetics, epigenetics and the environment work together to drive obesity have set for themselves a giant task. But it's one they must accomplish if they hope to have an impact on the obesity epidemic. "A proper understanding of the biology is going to help us determine what can be done," Bouchard says. "If we don't have that, we're shooting in the dark." ■

Cassandra Willyard *is a freelance science writer based in Madison, Wisconsin.*

1. Montague, C. T. *et al. Nature* **387**, 903–908 (1997).
2. Farooqi, I. S. & O'Rahilly, S. *Arch. Dis. Child.* **83**, 31–34 (2000).
3. Frayling, T. M. *et al. Science* **316**, 889–894 (2007).
4. Karra, E. *et al. J. Clin. Invest.* **123**, 3539–3551 (2013).
5. Smemo, S. *et al. Nature* **507**, 371–375 (2014).
6. Kilpeläinen, T. O. *et al. PLoS Med.* **8**, e1001116 (2011).
7. Pearce, L. R. *et al. Cell* **155**, 765–777 (2013).
8. Farooqi, I. S. *et al. J. Clin. Invest.* **110**, 1093–1103 (2002).
9. Lubrano-Berthelier, C. *et al. Ann. NY Acad. Sci.* **994**, 49–57 (2003).
10. Farooqi, I. S. *et al. N. Engl. J. Med.* **348**, 1085–1095 (2003).
11. Ravelli, G.-P., Stein, Z. A. & Susser, M. W. *N. Engl. J. Med.* **295**, 349–353 (1976).
12. Heijmans, B. T. *et al. Proc. Natl Acad. Sci. USA* **105**, 17046–17049 (2008).
13. Ng, S.-F. *et al. Nature* **467**, 963–966 (2010).
14. Soubry, A. *et al. BMC Med.* **11**, 29 (2013).
15. Hackett, J. A. *et al. Science* **339**, 448–452 (2013).
16. Skinner, M. K. *et al. BMC Med.* **11**, 228 (2013).

© 2014 Macmillan Publishers Limited. All rights reserved



CENTRE FOR INFECTIONS/PUBLIC HEALTH ENGLAND/SCIENCE PHOTO LIBRARY

Some of the roughly 1,000 bacterial species in the human gut help make us fat, while others keep us lean.

MICROBIOME

A complicated relationship status

Nothing is simple about the links between the bacteria living in our guts and obesity.

BY SARAH DEWEERDT

A few years ago, Liping Zhao, a microbiologist at Shanghai Jiao Tong University in China, put a man with a body mass index of 58.8 — classified as very severely obese — on a strict diet. Over the course of 6 months, the man shed more than 50 kg. In addition, a group of bacteria known as *Enterobacter* became undetectable in his stool samples, even though they had previously made up 35% of the microbes in his gut[1].

The decline and fall of a set of bacteria might seem incidental to the man's impressive weight loss, but Zhao and many other researchers say that the human gut microbiota — the assortment of 1,000 or so species of bacteria that inhabit our digestive tract — has an important role in regulating body weight.

"It's not calories alone," Zhao says, that determine whether a person is obese. To keep the weight down, "You also need to take care of the nutritional needs of beneficial bacteria in the lower gut." Similarly, some components of a healthy diet may curtail the growth of obesity-promoting bacterial strains.

Researchers are still unravelling the relationship between diet, gut microbes and body weight. "There are a lot of studies in humans, but those are only associations. There are a lot of studies of causation, but those are only in animals," says Fredrik Bäckhed, a researcher at the University of Gothenburg in Sweden who investigates the gut microbiota using mouse models.

The task now, say Bäckhed and others, is to translate results from studies of lab mice into treatments for humans in the real world.

That is far from straightforward. Last year, Zhao conducted a clinical trial of the dietary regimen that caused the dramatic weight loss in his severely obese subject, including whole grains, traditional Chinese medicinal foods and 'prebiotics' — supplements that promote the growth of beneficial gut microbes. After 9 weeks, the nearly 100 study participants had improved markers of metabolic health and lower levels of potentially harmful bacteria, including *Enterobacter*, but they only achieved a modest weight loss of about 6 kg on average[2].

But clinical trials into microbe-based interventions are just getting started, which is not surprising given the fact that serious research connecting gut microbes to obesity began scarcely a decade ago.

GUT REACTION

The first clues to this relationship came from mice that lack the gene for leptin, a hormone that regulates appetite. These mice eat insatiably, and as adults typically weigh three times as much as normal mice.

In 2005, researchers led by Jeffrey Gordon at Washington University in St Louis, Missouri, reported that the gut microbiota of leptin-deficient mice contains about 50% fewer Bacteroidetes and 50% more Firmicutes, each a major group of bacteria, compared with normal mice[3]. "This was the first direct evidence that

© 2014 Macmillan Publishers Limited. All rights reserved

there were differences in the microbial communities between lean and obese mammals," says Robin Knight, a computational biologist at the University of Colorado, Boulder, who collaborated on the study.

The following year, researchers in Gordon's lab identified the same pattern in humans: obese individuals have fewer Bacteroidetes and more Firmicutes than lean people. Moreover, the proportion of Bacteroidetes increases when individuals lose weight[4]. In fact, shifts in the balance of these two types of bacteria crop up again and again in research into the gut microbiota, diet and obesity.

As DNA sequencing becomes faster and cheaper, researchers have begun to look at gut microbes in finer detail. These days, they are analysing not just the kinds of microorganisms present in the gut, but also the genes that those microorganisms carry. (Scientists generally refer to the collection of bacterial species present in the gut as the microbiota, and the collection of genes as the microbiome.)

In 2013, a group of researchers collaborated with the MetaHIT Consortium, a European effort to determine the associations between gut microbes and chronic diseases, to sequence the microbiomes of 169 obese and 123 non-obese individuals[5]. They found that people fell into two groups that differed in the diversity of the microbial genes represented in their guts. Those with fewer genes tended to have more body fat and other markers of poor metabolic health compared with people with a more diverse microbiome.

Similarly, Knight and his colleagues calculated that microbial genes sort the lean from the obese with 90% accuracy, whereas looking at human genes yields the right answer only 58% of the time. "Microbial genes are a much better readout of whether you're likely to be obese or not than human genes are," Knight says.

CAUSE OR EFFECT?

But just showing that people have different microbes is not enough. "You don't necessarily know whether the microbial changes are a cause or an effect of the obesity," Knight points out.

So another line of research is aimed at establishing causality. "The strongest pieces of evidence are the mouse studies that have been done where the microbiota of an obese mouse is transferred to a germ-free mouse," says Rosa Krajmalnik-Brown, a microbiologist at Arizona State University in Tempe.

Germ-free mice, which lack gut microbes altogether because they are delivered by Caesarean section and raised in special aseptic cages, have lower body fat than conventionally raised mice. Gordon and his colleagues have found that when a germ-free mouse is colonized with gut microbes from a normal mouse, it experiences a 60% increase in body fat over the course of 2 weeks — despite eating less food than it did before the transfer[6].

➲ NATURE.COM
Impact of dietary intervention on microbial genes:
go.nature.com/jwmqho

"That provided the first mechanistic evidence that something about the microbes in our gut is increasing our ability to store body fat," says Peter Turnbaugh, a systems biologist at Harvard University in Cambridge, Massachusetts, who worked as a postdoctoral researcher in the Gordon lab.

Furthermore, the microbiomes of obese individuals have a different effect than those of normal-weight mice. "They gain about twice as much body fat over the course of two weeks if you colonize them with a sample that comes from an obese donor," says Turnbaugh. "And that can be from a mouse that's obese because of a genetic mutation in leptin, or mice that are obese due to consuming a high-fat, high-sugar diet." Researchers have even shown that germ-free mice that receive gut microbes from an obese human donor gain more weight than those that receive them from a lean person[7].

But not everyone finds these data convincing. Germ-free mice given obesity-associated microbiota gain weight, but they do not actually become obese themselves, points out Eric Martens, a microbiologist at the University of Michigan Medical School in Ann Arbor. "The magnitude of the change never really comes back to anything above what a normal mouse would have," Martens says. "You're not looking at transplantable obesity."

Moreover, diet is a major factor in obesity, and diet also shapes the microbiota. Often, changes in the levels of gut microbes produced by healthy or unhealthy diets are broadly similar to the differences seen in lean versus obese individuals. "So you already have this diet-to-microbiota relationship that's difficult to disentangle from the microbiota-to-obesity relationship," Martens says.

Others say the tight coupling of diet and the microbiota is the point. Zhao has shown that mice colonized with *Enterobacter cloacae* B29, a bacterial strain isolated from his obese patient, become obese themselves if they are fed a high-fat diet, but not if they are fed a normal diet. "As microbiologists we know for a pathogen to cause a disease you need many things," Zhao says. "First you need the pathogen, but then you also need the right environmental condition to trigger the problem."

Demonstrating similar causality in humans will require additional work, however. "What we really need is prospective studies where we see an altered microbiome before the disease onset," says Bäckhed. Several groups are now beginning these investigations.

A BUG'S LIFE

Meanwhile, these questions have not stopped researchers from starting to look at microbe-based approaches to treating obesity. But for this to be more than educated guesswork, scientists will have to figure out the precise molecular and biochemical mechanisms that link diet, gut microbes and body weight. They will also need to identify the particular bacteria, at species level, that may be involved.

Early work in this area suggested that the genes and biochemical pathways characteristic of obesity-associated microbiomes are more efficient at extracting energy from food than are those of normal-weight individuals. Essentially, an obese mouse gets more calories out of a cup of mouse chow than a lean mouse does. "Our hypothesis at that time was basically about energy harvest, that the microbiome helps to digest carbohydrates that would otherwise be indigestible to the host," recalls Bäckhed, who worked on these questions as a postdoc in the Gordon lab.

But the picture soon got more complicated, as Bäckhed and others found that germ-free mice do not become obese when fed a Western-style diet high in fat and simple sugars — components of food that gut microbes have only a minor role in digesting. "The germ-free gut was still protecting these mice against obesity, suggesting that there are other mechanisms at play as well," says Bäckhed.

"Microbial genes are a much better readout of whether you're likely to be obese or not than human genes are."

Researchers have since shown that the gut microbiota can affect the body's signalling systems related to hunger and feeling full[8], and even how quickly food passes through the gut[9]. Other studies have traced how diet and microbes can interact to produce inflammation and an impaired gut barrier, or 'leaky gut', which may contribute to obesity[10].

In mice, it is possible to reverse many of these effects with prebiotics[11], the most widely used of which is oligofructose, a type of indigestible carbohydrate found in foods such as bananas, garlic and Jerusalem artichokes. "We found that mice fed with oligofructose had an improved gut barrier function," says Patrice Cani, a researcher into metabolism and nutrition at the Catholic University of Louvain in Belgium. The mice that were given prebiotics also had improved metabolic markers, reduced fat mass and reduced inflammation, Cani adds.

OF MICROBES AND HUMANS

It is not clear, however, how well these outcomes translate to humans. Last year, Cani and his colleagues reported that obese women who took a supplement of oligofructose and a similar substance called inulin every day for three months showed a slight decrease in fat mass and a reduction in blood levels of an inflammation-promoting molecule[12]. But the results "were not really equivalent to the ones we observed in mice", Cani says.

Prebiotics are only one of several strategies to manipulate the gut microbiota. Other

© 2014 Macmillan Publishers Limited. All rights reserved

LEAN OPERATION

Does the microbiota determine the success of gastric surgery?



VISUALS UNLIMITED/GETTY IMAGES

One of the most drastic approaches to weight loss in obese individuals is gastric bypass surgery. Surgeons make the stomach smaller and reroute the small intestine, reducing the body's ability to absorb nutrients from food.

It is a profound reorganization of the digestive system. Yet recent studies suggest that a major portion of the weight loss triggered by bypasses — commonly 4.5–9.0 kg per month for the first year — may be attributable to changes in the gut microbiota that occur after surgery.

A study of three normal-weight individuals, three obese individuals and three people who had undergone gastric bypasses provided the first evidence for this idea. "We analysed the bacterial communities in the three groups, and they were really different," says study co-leader Bruce Rittmann, a professor at the Biodesign Institute at Arizona State University in Tempe. "Each community was relatively similar within the group, but the groups were very different from each other."

Despite the study's small size, the differences were statistically significant. "Patients with a gastric bypass had a higher population of Gammaproteobacteria," says Rosa Krajmalnik-Brown, a microbiologist at Arizona State University and the other co-leader. This large group of microbes includes, for example, the familiar gut bacterium *Escherichia coli*. "That was one of the most striking differences."

It is not yet clear why these shifts occur, but researchers hypothesize that the changes in gut anatomy, together with different dietary habits post-surgery, may favour the growth of different bacteria in the gut. Bolstering the observations in humans, similar changes have been found in the microbiota of rats and mice that undergo gastric bypass operations. In fact, if gut microbes from a post-surgery mouse are transferred to a mouse lacking gut microbes, the recipient mouse loses weight[13].

That is particularly striking because such mice, without any gut microbes to help them extract energy from food, start off with a lower than normal percentage of body fat. "It's surprising that given the right configuration of microbiota you can actually drive their weight even lower," says study team member Peter Turnbaugh, a systems biologist at Harvard University who works on gut microbes.

But some individuals who undergo gastric bypass surgery regain much of their weight after a few years. In an ongoing study, Rittmann and Krajmalnik-Brown are analysing the gut microbiota of up to 40 individuals several years after gastric bypass surgery to try to identify differences in the microbiota of successful and unsuccessful patients.

The team also plans to follow about ten new gastric bypass patients to track the changes in their gut microbiota after the surgery. "We're trying to find a biomarker," Krajmalnik-Brown says. "That would be really great if we could increase the success rate of surgery just by knowing the characteristics of the gut microbiota that are linked to success." Perhaps one day these patterns will yield insights into how to manipulate the microbiota to produce weight loss even without surgery. ***S.D.W.***

possibilities include faecal transplantation — in which communities of bacteria from one individual are given to another — or consuming beneficial bacteria as probiotics.

But so far, these other approaches have also worked better in mice than in humans. For example, a small faecal transplantation study found that people who received microbiota from a lean donor experienced improvements in insulin sensitivity, but no change in body mass index.

The top candidate for a bacterial species that could be given as a prebiotic to decrease body weight is *Akkermansia muciniphila*. This bacterium is found in most people's guts, but at lower numbers in obese individuals. When people lose weight after gastric bypass surgery (see 'Lean operation'), *A. muciniphila* flourish.

Mouse studies suggest that *A. muciniphila* strongly affects body weight. "If you take a normal mouse and give it a high-fat diet, it becomes obese. That's not rocket science," says Willem de Vos, a microbiologist at Wageningen University in the Netherlands, who was part of the team that discovered the species in 2004. "But if you give a high-fat diet and you give *Akkermansia* at the same time, there's no obesity."

Now, de Vos and his collaborators are beginning a clinical trial in which obese patients will be given *A. muciniphila* to see if the bacterium can help them to lose weight.

It is probably too soon to get excited. Even if the science is straightened out, it will be hard to develop an effective microbiota-based weight intervention for humans. Lab mice, after all, live tightly controlled lives, whereas we are constantly surrounded by temptation. "It's so easy to be a nibbler," as Bäckhed puts it.

People are also a lot more genetically diverse than lab mice. As a result, the effectiveness of diet and exercise interventions for obesity varies greatly between individuals, and the same is likely to be true for microbe-based treatments. "If you put two people on the same diet they're not going to come out with the same intestinal microbiota," says Krajmalnik-Brown. A person's initial microbiota, metabolism and even their gut anatomy may influence the results of a prebiotic or probiotic regimen.

"Lab mice live tightly controlled lives, whereas we are constantly surrounded by temptation."

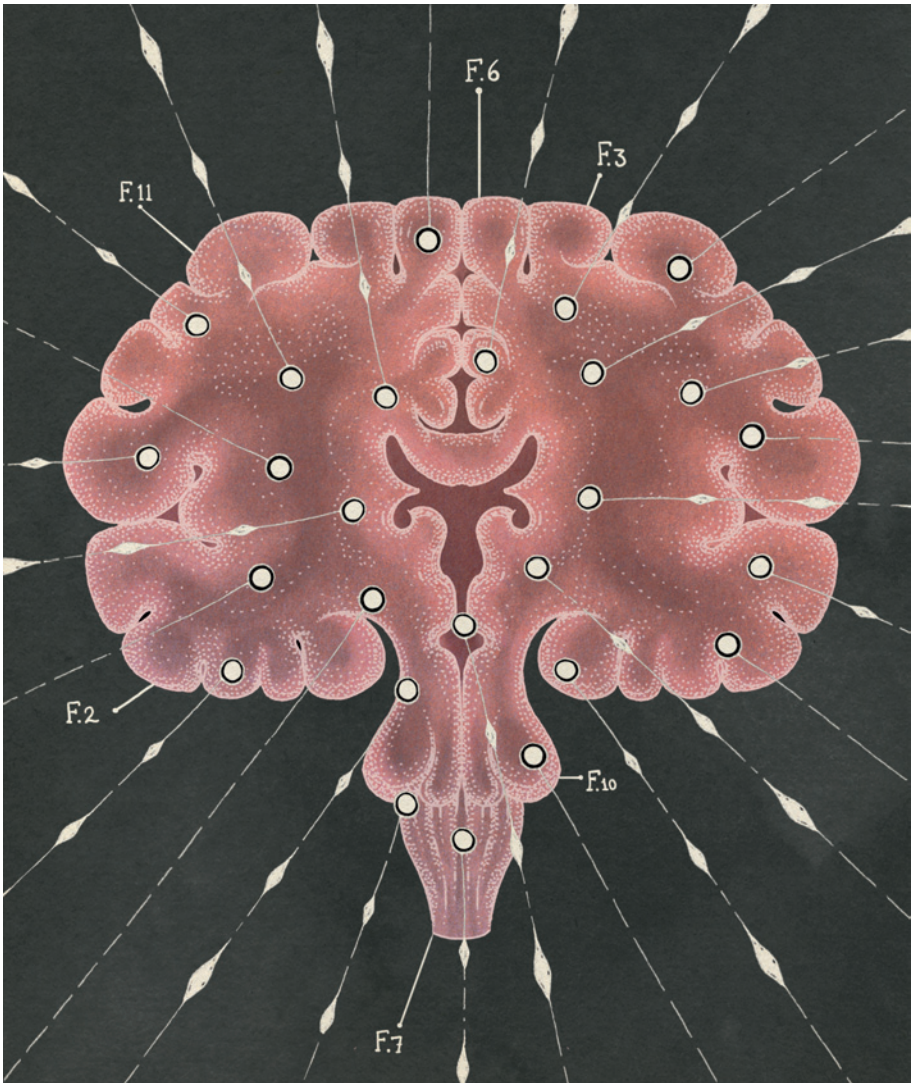
It may eventually be possible to analyse a person's existing microbiota to predict the effectiveness of various treatments. Already scientists have shown that individuals with a gene-rich microbiome respond differently to diets than those with fewer genes in their gut.

But that is still far away. "What we really don't know in humans is the question that is most relevant for a person who is obese, which is how much of your body fat is really caused by gut microbes," says Turnbaugh. "And that might vary a lot from person to person." So microbes may punch above their size, but the relationship between them and us is what finally matters, and that remains, well, complicated. ■

Sarah DeWeerdt *is a freelance science writer based in Seattle, Washington.*

1. Fei, N. & Zhao, L. *ISME J.* **7**, 880–884 (2013).
2. Xiao, S. *et al. FEMS Microbial Ecol.* **87**, 357–367 (2014).
3. Ley, R. E. *et al. Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
4. Ley, R. E. *et al. Nature* **444**, 1022–1023 (2006).
5. Le Chatelier, E. *et al. Nature* **500**, 541–549 (2013).
6. Bäckhed, F. *et al. Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
7. Ridaura, V. K. *et al. Science* **34**, 1241214 (2013).
8. Schéle, E. *et al. Endocrinology* **154**, 3543–3551 (2013).
9. Wichmann, A. *et al. Cell Host Microbe* **14**, 582–590 (2013).
10. Cani, P. D. *et al. Diabetes* **56**, 1761–1772 (2007).
11. Everard, A. *et al. Diabetes* **60**, 2775–2786 (2011).
12. Dewulf, E. M. *et al. Gut* **62**, 1112–1121 (2013).
13. Liou, A. P. *et al. Sci. Transl. Med.* **27**, 178ra41 (2013).

© 2014 Macmillan Publishers Limited. All rights reserved



KATIE SCOTT

NEUROSCIENCE

Dissecting appetite

A slew of new technologies are helping to map the neural circuits that control when, and how much, we eat.

BY BIJAL P. TRIVEDI

Many people think that overeating results from flagging willpower. But feeding behaviour is actually under intense biological control. Bradford Lowell, a neuroscientist now at Beth Israel Deaconess Medical Center in Boston, Massachusetts, realized this in the late 1970s when, as an undergraduate at the University of Massachusetts, he made tiny nicks in the brains of rats to see which regions might be responsible for the obesity observed in earlier brain lesion studies. As the animals began to overeat and became massively obese, Lowell "was amazed that such tiny lesions could have such an enormous effect."

Lowell had cut neurons near the paraventricular nucleus of the hypothalamus — the nerve centre of feeding behaviour. The hypothalamus comprises at least a dozen different nuclei (clusters of many different types of neuron), and receives and emits a variety of signals for both hunger and satiety. It also sets the body's circadian clock, and receives sensory data: sights, smells and tastes. Visceral cues from our stomach and intestines, and reward and motivational triggers from elsewhere in the brain, also converge here. "It's a tangle of circuits that looks like a Jackson Pollock painting," says Lowell.

A picture of the anatomical configuration of this complex web of neurons is of little value, however, if the function of a particular neuron, and where it fits in the wiring diagram, is unknown. In the 1980s, there were few ways of doing mechanistic studies in the brain other than making cuts. But in 1989, the development of knockout mice and related *Cre*/*lox* conditional knockout technologies made it possible to understand the function of a neuron. In the past few years, optogenetics and chemogenetics — which use light or designer drugs to switch off individual neurons — have made it feasible to map the neural networks that control eating.

UNDERSTANDING APPETITE

To tease apart the hunger circuits, researchers initially focused on a tiny section of the hypothalamus called the arcuate nucleus. "It's an amazing structure," explains Lowell. "It exerts powerful control over hunger." The region contains two key neurons: AgRP neurons, which sense a fasted state and stimulate appetite, and POMC neurons that inhibit appetite. Together, these neurons monitor the nutritional state of the body and whether it has a surplus or deficit of calories by sensing chemicals in the blood and signals from upstream neurons. But Lowell isn't alone in being fascinated by the power of neurons to govern appetite.

The AgRP neurons were also being studied by Richard Palmiter, a neurobiologist at the University of Washington in Seattle. Palmiter was drawn to the neural circuitry controlling eating when his student made a mouse lacking the gene needed to produce neuropeptide Y (NPY), one of the appetite stimulators produced by AgRP neurons. When NPY is injected into rodent brains, the animals eat voraciously. Palmiter expected mice without NPY to ignore food and lose weight, yet the NPY knockout mice fed normally and followed a normal growth curve[1]. Researchers at Merck Research Laboratories in Rahway, New Jersey, found the same thing when they knocked out both NPY and another appetite stimulator, Agouti-related peptide (AgRP), also produced by AgRP neurons[2].

Were AgRP neurons unimportant in appetite regulation? Or did they secrete so many appetite-stimulating neurotransmitters that knocking them out one at a time had no effect? To find out, Palmiter destroyed the AgRP neurons entirely. Over six days, the mice ate less and lost 20% of their body weight. "We were shocked," says Palmiter. "These neurons were vital. If we got rid of them the animals would starve to death." It seems the AgRP neurons blocked a neural pathway that otherwise suppressed appetite and mediated anorexia.

© 2014 Macmillan Publishers Limited. All rights reserved

Apart from NPY and AgRP, what other appetite-promoting signals did these neurons produce? Palmiter had a hunch that the missing piece of the puzzle was GABA, which the AgRP neurons also secrete. He destroyed the AgRP neurons again, but this time he also stimulated GABA receptors in the parabrachial nucleus — a relay station that receives sensory inputs, including those from the tongue and gut, and sends signals to other brain regions. After 10 days of GABA signalling, the mice began to eat again, even without the AgRP neurons. Mammals seemed to have evolved a mechanism to prevent starvation when AgRP neurons were destroyed[3].

CIRCUITS AND SWITCHES

To control feeding, the brain needs to constantly monitor the body's nutritional status and energy needs — and the neurons in the ARC initiate this behaviour. Scott Sternson, a neuroscientist at the Howard Hughes Medical Institute's Janelia Farm campus in Ashburn, Virginia, has been studying the AgRP and POMC neurons, which promote and suppress eating, respectively. Using optogenetics, he showed that activating only 800 AgRP neurons gave the mice a voracious appetite. "What fascinates me," says Sternson, "is that we can activate this tiny group of AgRP neurons and trigger very complex behaviour — not just chewing and swallowing, but obsessive searching and other complex behaviours to get food."

Controlling feeding behaviour is like using the accelerator in a car, explains Sternson. The more AgRP neurons are excited, the stronger the desire to eat. That probably suggests that as we get hungry, more nutritional signals circulate in the blood and activate increasing numbers of AgRP neurons.

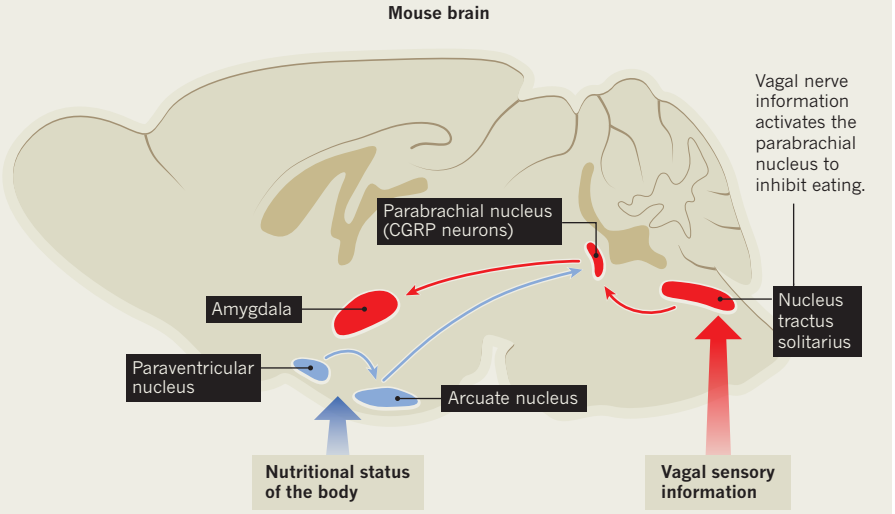
Sternson's team activated POMC neurons at the beginning of the dark cycle when mice typically feed, but this didn't initially affect food intake. Over the next 24 hours, however, the mice ate dramatically less food and lost 7% of their body weight. This suggests that AgRP and POMC neurons together regulate feeding behaviour and body weight over longer timescales, from hours to days[4].

Lowell used chemogenetics to investigate the reverse scenario — activating AgRP neurons during the day, when mice do not normally feed. These mice searched for food as if they were starving[5]. When neurons were activated twice a day, the mice gained weight; when the neurons were no longer activated, the mice stopped overeating and returned to their normal weight. "The AgRP neurons greatly increase the reward value of food — causing the seeking and consumption of food," says Lowell.

But neurons do not work in isolation. The next step was to identify which neurons triggered the AgRP neurons. By using a rabies virus expressing a reporter gene that makes active upstream neurons glow green, Lowell traced the AgRP neural circuit back to the hypothalamic paraventricular nucleus[6] — the region associated with satiety that Lowell had cut decades earlier.

THE HUNGER CIRCUITS

The neural circuits and chemicals that make us seek out food and eat, and those that stop us, are being charted in the brains of mice. Neuroscientists expect there will be analogous circuits in the human brain that could potentially be manipulated with drugs to treat obesity or loss of appetite.



Lowell showed that activating a tiny subset of neurons in this part of the hypothalamus — those that express the proteins TRH and PACAP — would activate the AgRP neurons and drive well-fed mice to eat voraciously. Conversely, if these neurons were turned off, starving mice would barely eat.

From his experiments with AgRP, NPY and GABA, Palmiter knew that without signals from the AgRP neurons, the parabrachial nucleus would inhibit feeding. But he didn't know which neurons were responsible for the loss of appetite. He thought that without input from AgRP neurons, calcitonin gene-related protein (CGRP) neurons in the parabrachial nucleus would become overactive, leading to starvation.

To test this hypothesis, Palmiter created mice with CGRP neurons that responded to blue light, and illuminated these neurons with blue laser light from optical fibres in the skull. Turning on the laser activated the CGRP neurons — mice that should have been hungry after sleeping most of the day lost all interest in food. It was as if they got a signal that said "I'm not hungry after all," says Palmiter. "Flicking off the laser made the mice hungry again." When these neurons were repeatedly stimulated at 12-hour intervals for several days, the mice ate less and lost weight, indicating that stimulating CGRP neurons is sufficient to cause starvation[7].

To map neurons downstream of the CGRP neurons, Palmiter used a fluorescent protein to light up the axons. This revealed that they projected into the amygdala — a brain structure involved in memory and emotion. Palmiter says this could explain why we lose our appetites when we are nauseous or in pain — when eating might be harmful — or when we have an emotional experience or suffer food poisoning.

All this research — figuring out the appetite circuits and signalling molecules — offers tantalizing clues that it may be possible to design drugs to tackle eating disorders. For example, oxytocin neurons (OXT) in the hypothalamus are a downstream target of the AgRP neurons. The loss of these OXT neurons is seen in Prader-Willi syndrome, a rare genetic disorder characterized by poor growth, intellectual impairment and a chronic, insatiable appetite that often leads to obesity. But stimulating both AgRP neurons and OXT neurons at the same time blocks AgRP-driven feeding, suggesting that the AgRP–OXT neural circuit might be a good mouse model for studying Prader-Willi[8]. Sternson adds that molecules that can regulate AgRP, OXT or POMC neurons would be promising candidates for treating overeating and anorexia.

Intervening too early in one of these pathways to inhibit eating and promote weight loss, or to combat loss of appetite, could have undesirable side effects, however. So finding the last neural link that inhibits or promotes eating is vital. But the neural maps confirm that eating is hardwired and tightly governed by biochemical and physiological mechanisms — a sign that it should be possible to design targeted obesity and anorexia drugs. ■

Bijal P. Trivedi *is a freelance science writer based in Washington DC.*

1. Erickson, J. C., Clegg, K. E. & Palmiter, R. D. *Nature* **381**, 415–418 (1996).
2. Ollmann, M. M. *et al. Science* **278**, 135–138 (1997).
3. Wu, Q., Boyle, M. P. & Palmiter, R. D. *Cell* **137**, 1225–1234 (2009).
4. Aponte, Y., Atasoy D. & Sternson, S. M. *Nature Neurosci.* **14**, 351–355 (2011).
5. Krashes, M. J. *et al. J. Clin. Invest.* **121**, 1424–1428 (2011).
6. Krashes, M. J. *et al. Nature* **507**, 238–242 (2014).
7. Carter, M. E. *et al. Nature* **503**, 111–114 (2013).
8. Atasoy, D. *et al. Nature* **488**, 172–177 (2012).

© 2014 Macmillan Publishers Limited. All rights reserved

PERSPECTIVE



Tricks of the trade

Processed foods that dilute protein content subvert our appetite control systems, say **Stephen J. Simpson** and **David Raubenheimer**.

The weight of evidence indicates that the primary driver of the obesity epidemic in recent decades has been increased intake, rather than decreased energy expenditure. Not surprisingly, much research has focused on the source of these extra calories, with vigorous debate over whether fats or carbohydrates (especially refined sugar) are the main culprits.

But this approach misses a fundamental point. The obesity problem is best understood not as the result of the overconsumption of a single macronutrient, but from a skewing of the proportion of each macronutrient in our diet — notably the dwindling quantity of protein in processed food products. The paucity of protein relative to fats and carbohydrates in processed foods drives the overconsumption of total energy as our bodies seek to maintain a target level of protein intake.

BALANCING ACT

An optimal diet includes a mixture of foods that provide dozens of different nutrients in appropriate amounts and in the right proportions. Consuming such a balanced diet is essential for the growth and maintenance of all organisms, and for reproductive fitness.

So it's no surprise that animals, including humans, have evolved exquisite appetite systems to control the consumption of proteins, carbohydrate and lipids, as well as micronutrients such as sodium and calcium[1]. These systems work independently to push the organism towards eating a well-proportioned diet: a lack of one nutrient will signal "eat more", while an excess of another will signal "eat less".

As long as the animal eats the foods that it evolved to consume, this balance is maintained. The trouble comes when it eats a diet with a disproportionate quantity of a particular macronutrient, either because of a lack of appropriate foods in the environment or because its appetite control systems have been fooled or subverted. This is now happening in humans, as we live in a foodscape scarcely recognizable to our forebears.

In humans, as in many other animals, the appetite prioritizes protein over carbohydrate or fat[2]. The evolutionary explanation is straightforward: eating too little protein compromises growth, development and reproduction.

Many processed food products are protein-poor but are engineered to taste like protein. Many people therefore eat far too much fat and carbohydrate in their attempt to ingest enough protein. In this way, engineered foods subvert the appetite control systems that should be helping to balance the consumption of macronutrients. The results are striking. In the United States, the typical diet saw a 0.8% decline in protein concentration between 1971 and 2006. During this same period, the consumption of calories from carbohydrates and fats increased by 8%, a trend reflected in the rising prevalence of obesity[3], but protein intake remained almost unchanged.

The substitution of carbohydrates and fats for protein is driven by economics. Food manufacturers have a financial incentive to replace protein with cheaper forms of calories, and to manipulate the sensory qualities of foods to disguise their lower protein content. This leads to savoury-flavoured food that makes us think we're eating protein when in reality it is loaded with carbohydrates and fats. The manufacturers of animal feed have similar incentives to add carbohydrates, resulting in fatter livestock (and pets). The higher cost of protein drives consumers to buy cheaper processed food loaded with fat and carbohydrates — an effect that disproportionately affects people on tighter budgets.

EVADING EVOLUTION

The range of processed food becoming available is evolving faster than our appetite control systems. Simple sugars and fat were rare in our ancestral environment and highly prized. This may be why modern humans tend to favour the fatty and sugary foods that now surround us, and would also account for the trend in agriculture and industrialization towards producing readily digested carbohydrates and fats.

> "MANY PEOPLE EAT FAR TOO MUCH FAT AND CARBOHYDRATE IN THEIR ATTEMPT TO CONSUME ENOUGH PROTEIN."

Populations accustomed to a high-protein diet, such as hunter-gatherers, may be most at risk in moving to a modern diet where protein is in short supply. If their physiology reflects their ancestral diet, they may have appetite systems that strive for an even higher protein intake than populations whose ancestors switched earlier to an agricultural lifestyle with a high-carbohydrate diet.

Much of the discussion of diets focuses on calorie counting and simplistic attempts to demonize elements of a normal diet. However, these approaches fail to capture the inherently multidimensional nature of nutrition. It is clear that the balance of nutrients — especially protein, fat and carbohydrate — has profound effects on many critical physiological functions, including appetite, energy intake, obesity, cardiometabolic health, ageing, immunity and the microbial ecology of the gut[1,4]. Teasing apart these interactions, and providing dietary interventions that can help alleviate their negative effects on health, requires a more integrated view of nutrition than is offered by the hegemony of the calorie or the search for a single culprit. ■

Stephen J. Simpson *is academic director of the University of Sydney's Charles Perkins Centre, which takes a multidisciplinary approach to obesity, diabetes and cardiovascular disease.* **David Raubenheimer** *is Leonard P. Ullman chair in nutritional ecology and nutrition theme leader at the Charles Perkins Centre.*
e-mails: stephen.simpson@sydney.edu.au; david.raubenheimer@sydney.edu.au

1. Simpson, S. J. & Raubenheimer, D. *The Nature of Nutrition: A Unifying Framework from Animal Adaptation to Human Obesity* (Princeton University Press, 2012).
2. Gosby, A. K. *et al. Obesity Rev.* **15**, 183–191 (2014).
3. Austin, G. L., Ogden, L. G. & Hill, J. O. *Am. J. Clin. Nutr.* **93**, 836–843 (2011).
4. Solon-Biet, S. M. *et al. Cell Metab.* **19**, 418–430 (2014).

© 2014 Macmillan Publishers Limited. All rights reserved

THE VERMILION MARKET

Where thoughts crystallize.



JACEY

BY PRESTON GRASSMANN

The crystals turned the sunlight into fragments of story across the market stalls. There were crystals that hung from strings or were set in place on ornate stands. Some even adorned the bodies of carvers, jewels that bevelled the fragments of poems and flash-stories. One woman wore a pair of haiku earrings; another, a bracelet that was carved in such detail that its planes cast more than a thousand phrases at once. Parents and children stood among them, pointing up at the fountain that would soon bring the off-world buyers to our city. I followed the gaze of a child as she stared up at the tower, waiting for the moment when it would light up and sway with the motion of their descent.

Like my parents, I am a carver of stories, born into a life at the brightest edge of visibility. I looked up at the spires of our seven towers, and remembered how my mother taught me to carve, when I first learned how light could refract into language and meaning. As the rooms were emptying for the annual market, these memories began to fill the empty space. Most of the children were content to live in the glass story-worlds of their carvings, but there was never a time that I didn't dream of what life might be like on other worlds. I wanted to see the Old Yards of Amethyst, where the starships and other machines were being built. I wanted to walk over the Ice Bridge of Cerulean to view the aurora veils of the northern sky.

"Wouldn't our stories offer more if we could experience those other worlds?" I had asked my mother, but she had only smiled and looked up at the spire. "Someday, you will experience them," she had said. "But for now, use that desire in what you create, turn it into light."

From my father's stall, I watched the platform descend. I could hear the sound of it sliding along the cable. When it reached the ground level, there was a thundering sound as the locks slid away. A moment later, the buyers emerged and began to flow down from the hill. They approached the crystal market in small groups. It was easy to tell which world they were from by the colours they wore and the style of their clothes. But many also had holograms of their home-worlds and solar systems spinning around them, orbiting over their heads like orreries.

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

I watched a man from Indigo break away from the group and walk through the stalls, ignoring the call of vendors. He finally stopped in front of our stall. I felt my heart racing as he bent over and began examining the crystals. I saw him lift one that I had known well, turning the phrases slowly in his hand. He smiled as the words opened up around him. I remembered how pleasant it had been to carve that story, as if the light, travelling all the way from the sun, had carried something of all the worlds in between. But those phrases seemed so much more surreal then, as I tried to imagine what he saw in them. He looked away and squinted at my father, saying: "She carves so beautifully … doesn't she?"

"Thank you," I said, before my father could answer.

The stranger's smile flickered for a moment and then brightened as he turned to stare at me. I could see a reflection of the glass fountain in his eyes.

"You wrote this?" he asked.

"With a little help," I said, pointing to my father.

"It was your story, Clarity …" my father said. Through the years, my father had spent long hours teaching me how to work at each angle, carving the planes to make the refractions as precise as they could be.

"It can't be as exciting as what you've seen out there," I said, looking up at the hologram. His home-world spun above him, swirls of white cloud over indigo. A single moon orbited the planet, shadowed with the chasm cities of World Engineers.

"Out there, it's all about merchant wars and political struggles and economic divisions. The stories of your world offer something more than all that … obfuscation. They bring …" he looked at me as if I was part of a bigger story, a prism carved not by time and memory, but by an image of what I was to him "… Clarity," he said, reading my name through the glass, saying it as if it was the word he had been looking for. His eyes seemed to turn the sun into words, luminous phrases set against the fountain. "Thank you for your stories," he said. It almost seemed gratuitous as he handed me the payment. "I will return again for your next crystal."

He held the crystal in his hand and read the words out loud for a moment before walking away. I watched the hologram world and its moon spinning above him as he went.

"Congratulations, Clarity," my father said. "You have another admirer."

I turned and saw the warmth in my father's eyes, and I knew what he meant.

"You remind me of your mother," he said. "She moves in you like light through a crystal." He took one of mother's stories from the stall and turned it in his fingers. The colours flared up around us. I could feel the warmth of its refractions against my skin.

I looked up at the fountain. It seemed to rise into the sun, a diamond-glass shard piercing its molten heart. "Someday, I'll go out there, like her. But for now, I have an idea for my next story," I said, watching the fountain sway as passengers returned back to their ships above Vermilion. For the first time, I was eager to carve my dream into glass. ■

Aside from occasional forays into Tokyo's theatre scene, **Preston Grassmann** *has been known to cover canvases in garish swathes of paint. His most recent work was published in* Caledonia Dreamin'*, and is forthcoming in* Daily Science Fiction.

© 2014 Macmillan Publishers Limited. All rights reserved